ignou
THE PEOPLE'S
UNIVERSITY

Indira Gandhi National Open University
School of Management Studies

BMP-001
Business Research

BMP-001 Business Research

BMP-001
Business Research

School of Management Studies
Indira Gandhi National Open University

PROGRAMME DESIGN COMMITTEE BBARL

Mr. Kumar Rajagopalan CEO, RAI	Mr. Upamanyu Bhattacharya, CEO, Fashion Global Retail	Mr. BVM Rao Head HR, Shopper's Stop Ltd.	**Faculty Members SOMS, IGNOU** Prof. M.S.S. Raju Prof. Nawal Kishor Prof. Sunil Kumar Gupta Prof. Rashmi Bansal Prof. Madhulika P. Sarkar Prof. Anupriya Pandey Dr. Rajendra Prasad Prof. Subodh Kesharwani
Mr. Ambarish Datta Head Technology Aided Learning Reliance Retail	Mr. Prithwish Bose Head HR Learning & Development Reliance Communication	Mr. Nimish Shah, Head Operations Lifestyle Stores	
Mr. Sumit Chandra Vice President Buying & Merchandising FMCG Aditya Birla Retail Ltd.	Dr. NVR Nathan, COO Everonn Skills Ltd.	Mr. Arun Vishwanath Head Training Raymond's Ltd.	
Mr. Lawrence Fernandes, Director, Retail Learning, RAI	Prof. Nawal Kishor SOMS, IGNOU	Prof. N.V. Narasimham SOMS, IGNOU	Dr. V. Venugopal Reddy Regional Director Cochin, IGNOU

PROGRAMME DESIGN COMMITTEE BBARL

Prof. M.S. Senam Raju Director, SOMS, IGNOU	Mr. Lawrence Fernandes Director – Academics, Retail Learning, RAI, Mumbai	Prof. Rama Mohan Rao Vice- Chairman, Andhra State Council of Higher Education, Andhra Pradesh
Dr. Ritesh Gupta Dean, FDDI Punjab	Mr. Gaurang Govind Retail Expert, Ghaziabad, U.P.	Ms. Shradha Kurup Head VM, Raymond Ltd. Mumbai.
Prof. Nawal Kishor SOMS, IGNOU	Prof. Madhu Tyagi SOMS, IGNOU	Prof. Subodh Kesharwani SOMS, IGNOU
Prof. Rashmi Bansal SOMS, IGNOU	Dr. Madhulika P. Sarkar SOMS, IGNOU	Prof. Anupriya Pandey Professor, SOMS, IGNOU
Dr. Rajendra Prasad Assistant Professor, SOMS, IGNOU	Mr. Shwet Kumar Manager, RAI, Mumbai	Mr. Manoj Sarkar Academic Consultant RAI, Mumbai

COURSE DEGISN & DEVELOPMENT COMMITTEE

Prof. Ms. Senam Raju Director , SOMS	Prof. Deepak Chawla IMI, New Delhi	Prof. Naved, DMS Aligarh Muslim University, UP
Prof. Sanjeev Mittal USMS GGSIPU, New Delhi	Prof. H.K. Dangi Commerce Deptt., DSE, Delhi	Dr. Sudhansh Sharma SOCIS, IGNOU
Prof. Nawal Kishor SOMS, IGNOU	Prof. Sunil Kumar Gupta SOMS, IGNOU	Prof. Rashmi Bansal SOMS, IGNOU
Prof. Madhulika P. Sarkar SOMS, IGNOU	Prof. Anupriya Pandey SOMS, IGNOU	Dr. Rajendra Prasad SOMS, IGNOU
		Prof. Subodh Kesharwani SOMS, IGNOU Convener & Course Editor

COURSE PREPARATION TEAM BMP - 001

Unit -1
Adapted from MCO-03 (Unit-1)
Prof. D. Obul Reddy
Former Professor of Commerce &
Osmania University, Hyderabad

Unit -2
Adapted from BPC-03 (Unit-2)
School of Social Science

Unit 3 and 4
Prof. Madhulika P. Sarkar
SOMS, IGNOU

Unit -5
Adapted from MCO-03 (Unit-3)
Dr. Ashok Gaba, SOVET, IGNOU
& ECO -07 (Unit-5), SOMS

Unit -6
Adapted from MCO-03 (Unit-5)
Dr. C.G. Naidu, IGNOU

Unit 7, 12 and 13
Prof. Subodh Kesharwani
SOMS, IGNOU

Unit 8 and 11
Prof. Sudhansh Sharma,
SOCIS, IGNOU

Unit -9
Adapted from MCO-03
(Unit-6 & Unit-7)
Dr. G. K. Mitra (Unit -6) and
Prof. M. S. Senam Raju (Unit-7)
Director, SOMS, IGNOU

Unit -10
Adapted from MCO-022
(unit-3) and
MMPC- 005 (Unit- 3) &
ECO-07 (Unit-10,11,12)
SOMS

Unit -14
Adapted from BANC-110 (Unit-8)
School of Social Science

Unit -15
Adapted from MCO-03 (Unit-19)
Prof. D. Obul Reddy
Osmania University, Hyderabad

COURSE EDITORS AND COORDINATOR

Prof. Subodh Kesharwani
SOMS, IGNOU, New Delhi

Prof. Madhulika P. Sarkar
SOMS, IGNOU, New Delhi

MATERIAL PRODUCTION

Mr. Tilak Raj
Assistant Registrar
MPDD, IGNOU, New Delhi

February, 2025

ISBN:

Further information on the Indira Gandhi National Open University courses may be obtained from the University's office at Maidan Garhi, New Delhi-110 068.

Printed and published on behalf of the Indira Gandhi National Open University, New Delhi, by the Registrar, MPDD, IGNOU.

Laser typeset by Tessa Media & Computers, C-206, A.F.E-II, Jamia Nagar, New Delhi

BLOCK 1
BUSINESS RESEARCH CONCEPTS

BLOCK 1 BUSINESS RESEARCH CONCEPTS

This block introduces the foundational concepts of business research. It provides an overview of research, including its objectives, classifications, significance, and the steps involved in the business research process. Additionally, it covers the importance of conducting a literature review, the "4 Cs" of literature review, key sources, and journal classifications. The block also explains the research problem, its significance, and the steps involved in identifying, selecting, and formulating research problems. Finally, it introduces the concept of research design, helping learners choose the most suitable design for their study.

The block titled "Business Research Concepts" consist of four units, as detailed below:

Unit 1: Business Research Process

This unit explains the meaning and definitions of research, along with its objectives and classifications. It outlines the characteristics of good research, its significance, and the steps involved in the business research process. While research is a systematic process, it is also dynamic and continuous, evolving over time.

Unit 2: Review of Literature

This unit focuses on the purpose, process, and importance of conducting a review of literature (ROL) in research. It explains how ROL helps identify research areas and provides insight into the "4 Cs" of literature review. The unit also discusses key sources of literature and the classification and indexing of journals, equipping learners to critically analyze and utilize academic resources effectively.

Unit 3: Research Problem

In this unit, learners are introduced to the concept of the research problem and its critical role in business research. It provides a step-by-step guide on identifying, selecting, and formulating research problems and objectives. To enhance clarity, this unit also includes case studies that demonstrate the application of these steps in real-world scenarios.

Unit 4: Research Design

This unit provides an overview of research design, its significance, and the various types of research designs used in the research process. Learners will understand the relationship between research problems and their corresponding designs, enabling them to differentiate between different types and select the most appropriate research design for their specific study.

UNIT 1 BUSINESS RESEARCH PROCESS

Structure

1.0 OBJECTIVES

After studying this unit, you will be able to:

- Understand the meaning of research;
- Comprehend the objective of research at different level;
- Analyse the Motivation and Characteristics of Effective Research;
- explore the role of research in business; and
- know the ethical issues in business research.

1.1 INTRODUCTION

Research in common parlance refers to search for knowledge. One can also define research as a scientific and systematic search for pertinent information on a specific topic. In fact, research is an art of scientific investigation. In other words, research is a part of any systematic knowledge. It has occupied the realm of human understanding in some form or the other from times immemorial. The thirst for new areas of knowledge and the human urge for solutions to the problems has developed a faculty for search and research and re-research in him/her. Research has now become an integral part of all the areas of human activity.

As we know that the business research process is a systematic way of gathering, analyzing, and interpreting data to create insights that can help guide an organisation's decision-making. If we talk, research from a business perspective, it is the process of gathering relevant information regarding a company's business activities and using it to maximise profit.

In this unit, we will discuss at length the need and significance of research, its types and process, the motivation behind doing it, and ethical concerns regarding it.

1.2 DEFINITION OF RESEARCH

The word Research is made up of two terms, **re** and **search; "re"** means "again", "a new", or "over" again, and "**search"** means to examine closely and carefully. Research is the careful, patient and systematic study in some field of knowledge undertaken to generalise facts. In simple words, research can be defined as the in-depth study of something in search of some new information.

The views of some scholars and sources on research are given below:

The systematic and objective analysis and recording of control observation that may lead to development of generalisation principles or theories resulting in predictions and possibility of utilisation ultimate control of events.

-John W. Best

Research is a scientific and systematic search for pertinent information on a specific topic.

-C. R. Kothari

An intensive and purposeful search for knowledge and understanding of social and psychological phenomena. It is not merely an accumulation of knowledge but a crucial and scientific analysis of social facts and formulation of generalisations as a basis of action and foresight.

-K. V. Rao

Research is generally oriented towards looking into **5 Ws and 1 H** of a problem, that is, Who, What, Where, When, Why, and How. As said by Bernard Baruch, 'Millions saw the apple fall but Newton was the one who asked why it was WHY'. It was this WHY which leads to the discovery of

the law of gravitation. So, a rational man does not accept any statement without empirical verification or logic. After the data/ facts have been collected, processed, analysed, we have to draw broad conclusions/ generalisations. Research provides an analytical framework for the subject matter of investigation. It establishes the relationship between the different variables. The cause-and-effect relationship between the different variables can also be identified, leading to valuable observations, generalisations and conclusions. Inductions and deductions are also possible in systematic research.

1.2.1 Inductive Research

Inductive research is a research method that involves collecting data, looking for patterns, and developing theories based on those patterns. It uses a "bottom-up" method in which the researcher starts with specific observations and then moves on to more general theories or ideas.

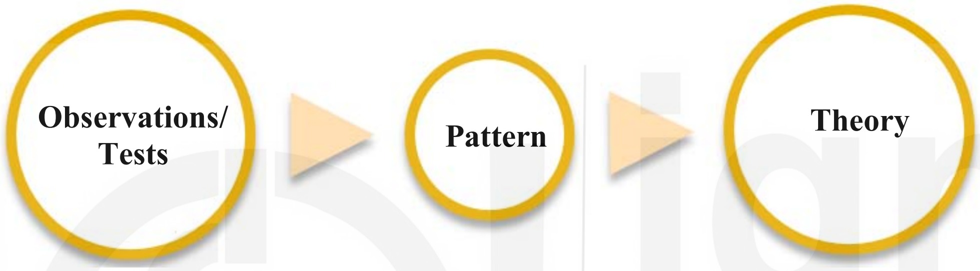


Figure 1.1: Inductive Research Process

These three steps are often repeated in a cycle, so the researcher can improve their analysis and understand the phenomenon over time. Inductive research aims to develop new theories and ideas based on the data rather than testing existing theories, as in deductive research. Inductive research is often used in exploratory studies or when not much research has been done on a topic before. *For example, if it is observed in a number of cases that when price increases less is purchased. Therefore, the generalization is "when price increases demand falls".* A researcher collects data on consumer behavior in a particular market and uses that data to identify trends and patterns that can inform marketing strategies.

1.2.2 Deductive Research

Deductive research is a type of research in which the researcher starts with a theory, hypothesis, or generalization and then tests it through observations and data collection. It uses a "top-down" method in which the researcher starts with a general idea and then tests it through specific observations. Deductive research is often used to confirm a theory or test a well-known hypothesis.

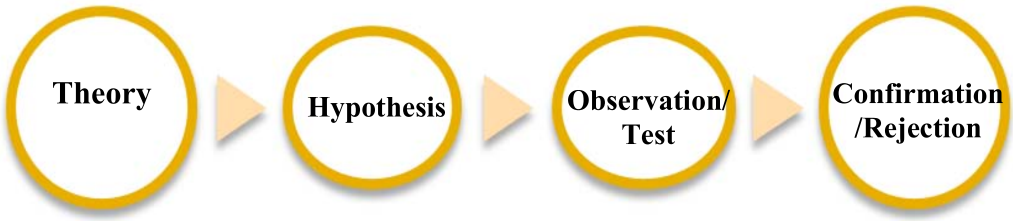


Figure 1.2: Deductive Research Process

The four steps of deductive research are repeated, and researchers may need to return to earlier steps if they find new information or new ways of looking at things. In contrast to inductive research, deductive research aims to test theories or hypotheses that have already been made. *For example, 'All men are mortal' is a general rule. Rakesh is a man. Therefore, from the general rule it can be deduced that Rakesh is also mortal'. In another example, All BBA degree holders are eligible for MBA and allied courses also is a general statement. Parth is a BBA degree holder. Therefore, it can be deduced that Parth is eligible for MBA.* Empirical studies have a great potential, for they lead to inductions and deductions. Research enables one to develop theories and principles, on the one hand, and to arrive at generalisations on the other. Both are aids to acquisition of knowledge.

1.2.3 Inductive Vs Deductive

The main differences between inductive and deductive research are how the research is done, the goal, and how the data is analysed. Inductive research is exploratory, flexible, and based on qualitative observation analysis. Deductive research, on the other hand, is about proving something and is structured and based on quantitative analysis.

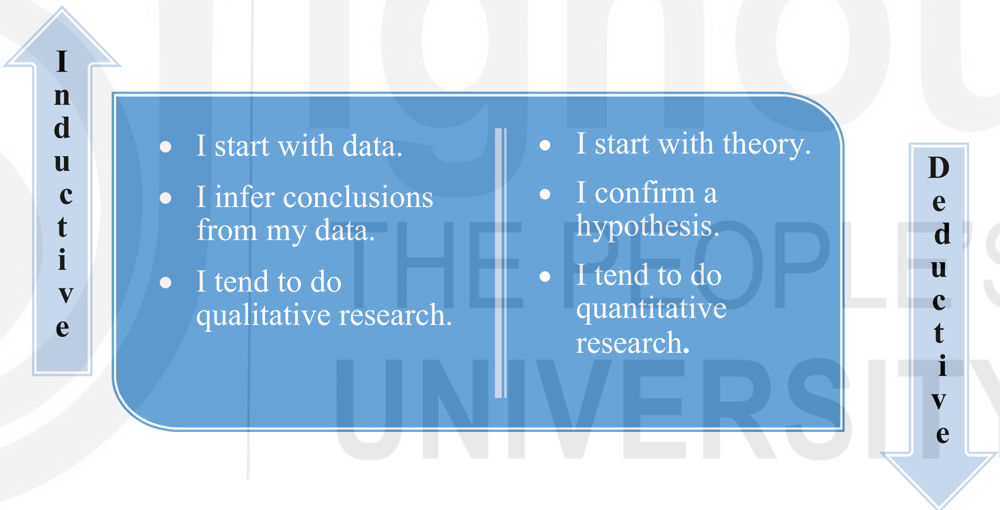


Figure 1.3: Inductive Research versus Deductive Research

Here are the main differences between inductive vs deductive research in more detail which you can able to understand clearly from the below table;

Table 1.1: Inductive vs. Deductive Research

Inductive Research	Deductive Research
• Bottom-up Approach	• Top-down Approach
• Develops theories from observations	• Tests theories through observations
• Used in exploratory studies	• Used in confirmatory studies
• Flexible and adapted to new findings	• Structure and systematic
• Relies more on qualitative analysis	• Relies more on quantitative analysis

From the above table this can clearly be observed that inductive and deductive research are two different approaches to research that differ in how they start and what they aim to do as inductive relies on patterns and trends while deductive relies on data and observations. But researchers can use both inductive and deductive approaches to gain a more comprehensive understanding of a topic such that new questions that arise during an investigation can be answered by using both approaches.

1.3 OBJECTIVES OF RESEARCH

Research uses scientific techniques to find answers to questions. The primary objective of research is to uncover undiscovered facts. However, so, each research project has a unique objective. When doing research, objectives can be defined at both the macro and micro levels. When we discuss research objectives at the macro and micro levels, we are distinguishing them between broader, overarching objectives and more specific, detailed objectives within a research study. Objectives at each level facilitate research in various ways.

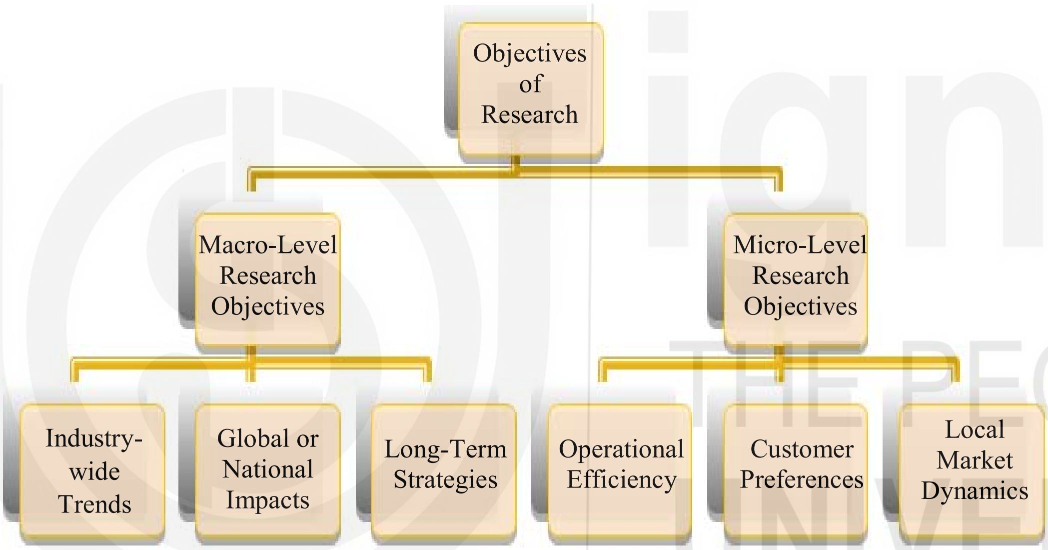


Figure 1.4: Objectives of Research

1.3.1 Macro-Level Research Objectives

Macro-level objectives focus on the broader aspects of the research. They aim to understand large-scale trends, patterns, or issues that have wide-reaching implications. These objectives might explore:

- **Industry-wide Trends:** Examining overall shifts in market demand, consumer behavior, or technological adoption across an entire industry.
- **Global or National Impacts:** Understanding the effects of economic policies, regulatory changes, or global events on a sector or economy.
- **Long-Term Strategies:** Identifying opportunities or threats that could influence the long-term direction of a business or industry.

For example: A study that aims to identify community needs and aspirations, determine broad educational goals, and develop suitable curricula.

1.3.2 Micro-Level Research Objectives

Micro-level objectives focus on specific characteristics within a broader context. These objectives typically deal with more focused, smaller issues that can impact day-to-day operations or individual elements of a business. These objectives might explore:

- **Operational Efficiency:** Investigating ways to optimize internal processes, reduce costs, or improve productivity within a specific department.
- **Customer Preferences:** Analyzing detailed customer feedback to refine product features, improve service quality, or tailor marketing strategies.
- **Local Market Dynamics:** Understanding the behavior of a specific target market or demographic within a particular geographic area.

For example: If your research is about understanding the impact of a new technology on retail businesses, a macro-level objective might be to assess how the technology is transforming the retail industry globally. In contrast, a micro-level objective might focus on how a specific retail store can implement this technology to improve customer experience and increase sales.

It is well versed that this dual approach ensures that the research addresses both the broader context and the specific, actionable insights necessary for practical implementation.

1.4 MOTIVATION IN RESEARCH

Motivation in research is the internal force that drives researchers to actively seek, continue with, and successfully conclude their research activities. The quality and impact of the research carried out are significantly influenced by this important factor. From the Figure 1.5 given below, learners can easily identify and understand the motivating factors that lead to research.

Figure 1.5: Motivation in Research

Researchers may be motivated by a variety of purposes, such as:

- **Personal Interest:** A genuine interest or passion for a certain subject area can motivate thorough research and continuous engagement.
- **Academic and Professional Goals:** Researchers may engage in intensive research activities driven by their desire to obtain degrees, gain academic reputation, or advancement in their professions.
- **Societal Impact:** A desire to address critical social issues or contribute to the improvement of societies might serve as a strong motivator.
- **Intellectual challenge:** It occurs once researchers are motivated to look into unfamiliar places in order to answer complex questions and expand the limits of present knowledge.
- **Recognition and Achievement:** The desire to be acknowledged for one's contributions to the area, whether through publications or prizes, can also operate as a driving force.

In simple terms, motivation in research acts as a basic framework that maintains the researcher's dedication and determination, ultimately contributing to significant developments and advancements in knowledge.

1.5 CLASSIFICATION OF RESEARCH

Research may be classified into different types for the sake of better understanding of the concept. Several bases can be adopted for the classification such as nature of data, branch of knowledge, extent of coverage, place of investigation, method employed, time frame and so on. Depending upon the basis adopted for the classification, research may be classified into a class or type. It is possible that a piece of research work can be classified under more than one type, hence there will be overlapping. It must be remembered that good research uses a number of types, methods, and techniques. Hence rigid classification of research is impossible. The following is only an attempt to classify research into different types.

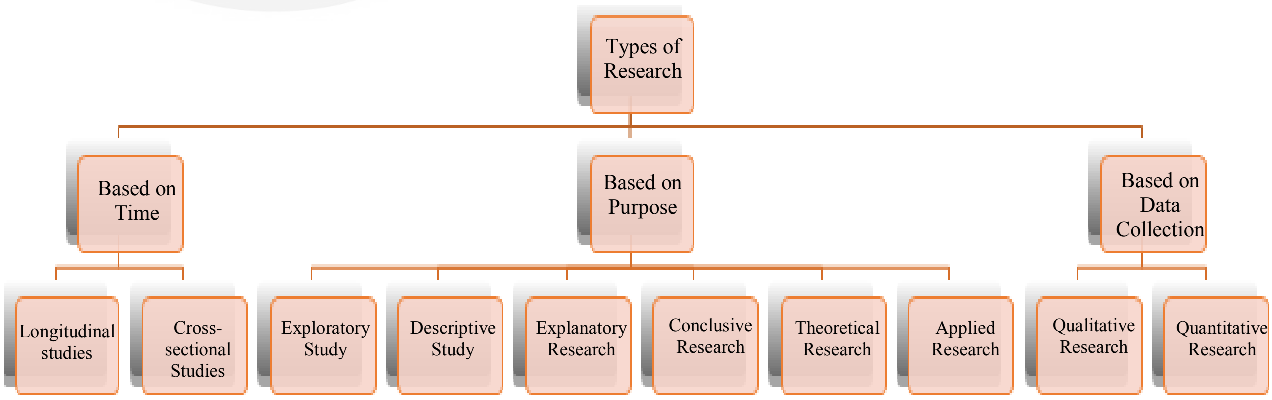


Figure 1.6: Types of Research

1.5.1 Based on Time

a) **Longitudinal studies (also referred to as Diachronic research):** It is the monitoring of the same event, individual, or group over a defined period of time. It aims to track changes in a number of variables and see

how they evolve over time. It is often used in medical, psychological, and social areas.

For example: A cohort study that analyses changes in a particular indigenous population over a period of 15 years.

b) **Cross-sectional Studies (also referred to as Synchronous Research):** Cross-sectional research design is used to observe phenomena, an individual or a group of research subjects at a given time.

For example: To validate the expected adoption rate and potential phone sales by enrolling men and women across regions and age ranges.

1.5.2 Based on Purpose

a) **Exploratory Study:** The information known on a particular subject matter is little. Hence, a study is conducted to know more about it so as to formulate the problem and procedures of the study. Such a study is called exploratory/ formulative study.

For example, an investigation of the role social media plays in the perception of self-image.

b) **Descriptive Study:** The major purpose of descriptive research is the description of a person, situation, institution or an event as it exists. Generally fact finding studies are of this type.

For example, investigating how the public census of influential government officials differs between urban and non-urban areas.

c) **Explanatory Research:** Explanatory research is the most common type of research method and is responsible for establishing cause-and-effect relationships that allow generalizations to be extended to similar realities. It is closely related to descriptive research, although it provides additional information about the observed object and its interactions with the environment.

For example, investigating the brittle behavior of a specific material when under compressive load.

d) **Conclusive Research:** Conclusive research is a type of research that provides information to help reach conclusions or make decisions. It's usually quantitative, meaning it uses numbers that can be summarised and quantified. Conclusive research relies on the use of highly structured techniques such as surveys with close-ended questions in order to prove or disprove a hypothesis.

e) **Theoretical Research:** Theoretical research also referred to as pure or basic research, focuses on generating knowledge, regardless of its practical application. Here, data collection is used to generate new general concepts for a better understanding of a particular field or to answer a theoretical research question.

f) **Applied Research:** The goal of applied research is to find strategies that can be used to address a specific research problem. Applied research draws on theory to generate practical scientific knowledge, and its use is very common in STEM fields such as engineering, computer science, and medicine.

1.5.3 Based on Data Collection

a) **Qualitative Research:** Qualitative research is often used in social sciences to collect, compare and interpret information, have a linguistic-semiotic basis, and are used in techniques such as discourse analysis, interviews, surveys, records, and participant observations.

In order to use statistical methods to validate their results, the observations collected must be evaluated numerically. Qualitative research, however, tends to be subjective, since not all data can be fully controlled. Therefore, this type of research design is better suited to extracting meaning from an event or phenomenon (the 'why') than its cause (the 'how').

For example, examining the effects of sleep deprivation on mood.

b) **Quantitative Research:** Quantitative research study delves into phenomena through quantitative data collection and using mathematical, statistical, and computer-aided tools to measure them. This allows generalized conclusions to be projected over time.

For example, conducting a computer simulation on vehicle strike impacts to collect quantitative data.

Here are the main differences between quantitative and qualitative research in a more precise way, which you can understand clearly from the below table:

Table 1.2: Difference between Quantitative and Qualitative Research

Particulars	Quantitative Research	Qualitative Research
Category	Objective	Subjective
Type of Reasoning	Deductive reasoning used to synthesise data	Inductive reasoning used to synthesise data
Focus	Concise and narrow	Complex and broad
Application	Tests theory	Develops theory
Basis of knowing	Cause and effect relationships	Meaning, discovery, correlation
Basic Element of Analysis	Numbers and statistical analysis	Words, narrative
Scope	Single reality that can be measured and generalised	Multiple realities that are continually changing with individual interpretation

Check Your Progress A

1) Distinguish between inductive and deductive logic.

....

....

....

....

2) What is the role of R & D in business?

....

....

....

....

3) How does research influence business decisions?

....

....

....

....

....

4) Distinguish between qualitative and quantitative research.

....

....

....

....

....

5) List the various types of studies according to the purpose of the study.

....

....

....

....

6) True or False

   a) Inductive research starts with a general theory and moves towards specific observations.

   b) Business research is not significant for decision-making in organizations.

   c) Ethical issues are an important consideration in business research.

   d) A good research study should have clearly defined objectives and systematic methodology.

   e) Research can only be classified based on time and cannot be classified based on purpose or data collection.

1.6 CHARACTERISTICS OF A GOOD RESEARCH

The characteristics of good research are essential to ensure that the study is credible, reliable, and adds value to the field of inquiry. The Figure 1.7 given below clearly shed light on the key characteristics of a good and successful research which are further explained below in detail.

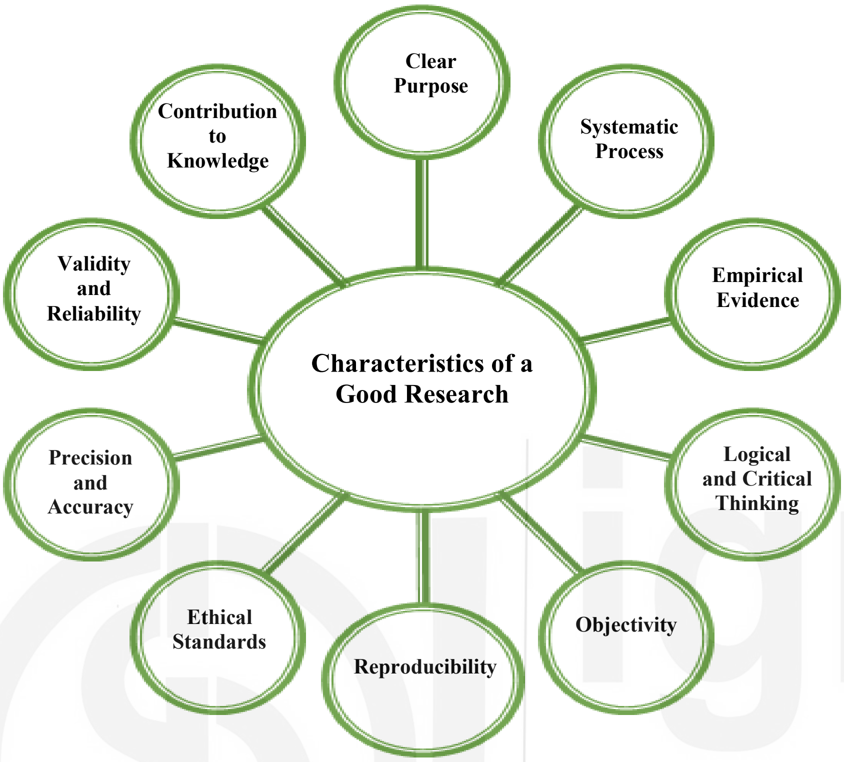


Figure 1.7: Characteristics of a Good Research

Here are key characteristics of a good research:

1. **Clear Purpose:** Good research starts with a well-defined purpose or objective, ensuring clarity in what the study aims to accomplish.
2. **Systematic Process:** It follows a structured approach, from problem identification to data collection, analysis, and conclusion, ensuring that each step is logically connected.
3. **Empirical Evidence:** Good research relies on factual, observable, and measurable evidence. Data is collected systematically and is subject to rigorous testing.
4. **Logical and Critical Thinking:** The research process is grounded in logical reasoning, where hypotheses or research questions are tested critically.
5. **Objectivity:** It should be unbiased, ensuring that personal opinions, emotions, or external pressures do not influence the outcomes.
6. **Reproducibility:** The research should be designed in a way that other researchers can replicate it under similar conditions, ensuring that the results are reliable and consistent.
7. **Ethical Standards:** Good research adheres to ethical guidelines, respecting participant consent, confidentiality, and data integrity.
8. **Precision and Accuracy:** The measurements, data analysis, and conclusions are accurate and precise, minimising errors and providing confidence in the findings.

9. **Validity and Reliability:** The study must measure what it claims to measure (validity) and produce consistent results over repeated trials (reliability).

10. **Contribution to Knowledge:** Ultimately, the research should contribute something new or valuable to the existing body of knowledge in its field, helping to inform future studies or applications.

These characteristics help in distinguishing high-quality research that is trustworthy and impactful.

1.7 SIGNIFICANCE OF BUSINESS RESEARCH

Business research is a never ending process of searching information on various factors that influences a business undertaking such as market forces, competitors, macro – economic environment, innovations and practices. Research information provided assists the management at different levels to make effective decisions so as to improve the performance of the firm. In detail, it can be described as a formal process through whose implementation information needed by an organization to make its decisions is obtained.

Figure 1.8: Business Research

Business Research is a formal and an organized method of inquiry used to investigate phenomena and make recommendations for business action. Helps companies to work with different business scenarios, with problems and to use opportunities in the most productive ways.

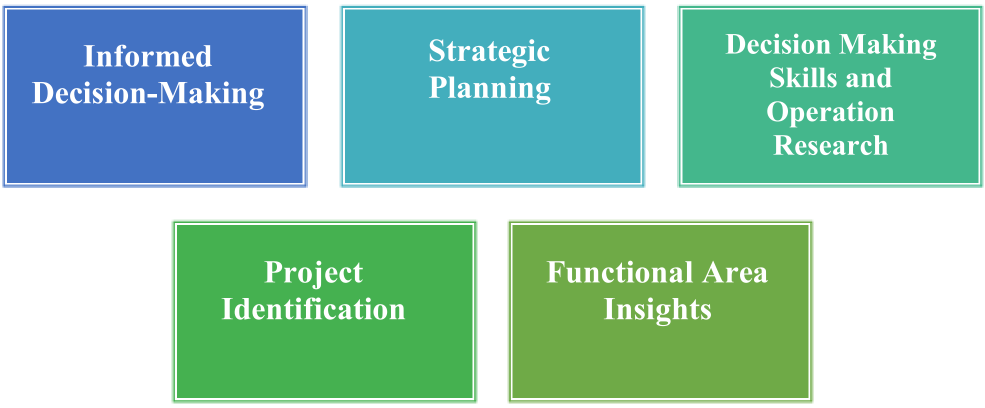


Figure 1.9: Significance of Business Research

The key roles and benefits of business research include:

a) **Informed Decision-Making:** Research cut out the trial and error and focuses on facts and thus the mishaps can be greatly minimized. It makes the decisions to be made timely, accurate, and efficient and in the best interest of the organization.

b) **Strategic Planning:** Strategic management is made easier through business research with data collected from the internal and external business environment used in planning, forecasting, and strategy formation.

c) **Decision Making Skills and Operation Research:** Research involves the use of quantitative techniques and reasoning to solve operational problems and find the best ways of handling dozens of managerial and administrative concerns.

d) **Project Identification:** Business research focuses on project identification and its feasibility. In addition, research helps in recognition of potential projects, their feasibility and management of the implementation processes.

e) **Functional Area Insights:** These are explained as follows:

   i) **Production:** Promotes improvement in product design, reduction in costs, increase in quality, and efficiency of the processes employed.

   ii) **Marketing:** Helps in providing market information for decision making, sales promotion and assessment of the impact of advertisement.

   iii) **Finance:** Enables management to control and make projections about the financial flows within the company.

   iv) **Human Resources:** Consults with policies concerning motivation, Job design, and the management of the workforce.

   v) **Adaptation to Change:** Concurrent research is important in helping the businesses to remain relevant and competitive by providing the relevant current trends, technologies and the market aspects.

   vi) **Innovation and R&D:** Essentially, research propels new product development, market identification, and effective methods of development leading to growth and development.

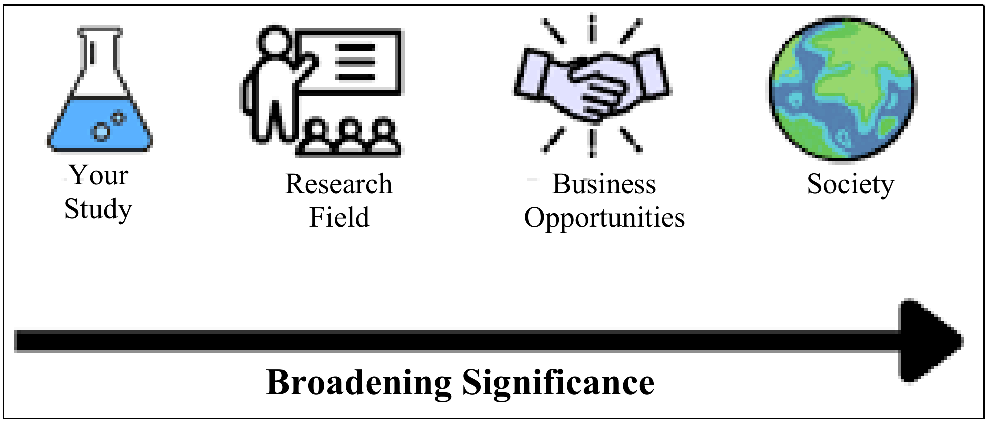


Figure 1.10: Broadening Research Significance

vii) **Interdisciplinary Connectivity:** As characterized by an interdisciplinary connectivity, which also refers to the extended integration of academic disciplines into an academic entity. Scholars have provided connections between different functional fields to support integrated decision making within the context of an organization.

Establishing research at all the organizational functionalities ensures companies thrive in competitive and volatile environments.

1.8 STEPS INVOLVED IN BUSINESS RESEARCH

One of the most crucial activities in the analysis of the aspects of business, market, competitors, and customers among others is business research. It plays a critical role in managerial decision-making, strategic management and fostering innovation in organizations and business entities. The procedures for business research may vary depending on the specific research objectives, but the most common methods are as follows. Let's delve into the various steps involved in business research:

Figure 1.11: Steps involved in Business Research Process

1. **Identifying the Problem or Opportunity:** All decision problems require a clear statement of the problem that has to be solved or the opportunity available to be taken. A well-defined problem is the foundation for effective research.

2. **Reviewing Literature:** The next step is to analyse the literature concerning the problem under study: it is crucial to reveal what has been research before.

3. **Formulating Research Objectives:** Set measurable objectives that outline how the research is supposed to get done to achieve the intended goals of study.

4. **Designing the Research Methodology:** Decide on the right approaches for data collection and data analysis Choose qualitative or quantitative methods based on the research objectives and goals.

5. **Collecting Data:** Implement the methods chosen for data collection, ensuring accuracy and reliability. Also follow the ethical considerations to maintain the integrity of the research process..

6. **Analyzing Data:** The subjects are used for interpretation of the findings following data collection, using statistical or thematic analysis, and generating meaningful conclusion that addresses the research objectives and knowledge database.

7. **Presenting Results:** Presently, our findings should be presented concisely and coherently in reports or presentations that outline suggestions and recommendations for particular stakeholders, as well comprehensively describing any challenges faced during the research.

8. **Drawing Conclusions:** Conclude the main discoveries of the study, and suggest how proposed understandings can be improved in order to enhance knowledge in the field.

9. **Disseminating Results:** Publish the results of the research by disseminating the findings in academic publications, conferences, and engagements within the community to contribute to the discourse and reach more people.

In summary, business research is defined as a formal process of formulating a business research problem and reviewing literature, business research purpose and objectives, identifying the right technique to conduct research and an appropriate approach in collecting data, analyzing data, and presenting and disseminating the findings which can be used. It provides a clear way from problem identification to gaining insights and knowledge in order to make relevant decisions in the business world.

1.9 ETHICAL ISSUES IN BUSINESS RESEARCH

Ethical considerations are important aspects of research design that can help ensure the safety of participants and guide research practices. But it is regrettable that ethical considerations are becoming inevitable in business research to maintain the authenticity of the research.

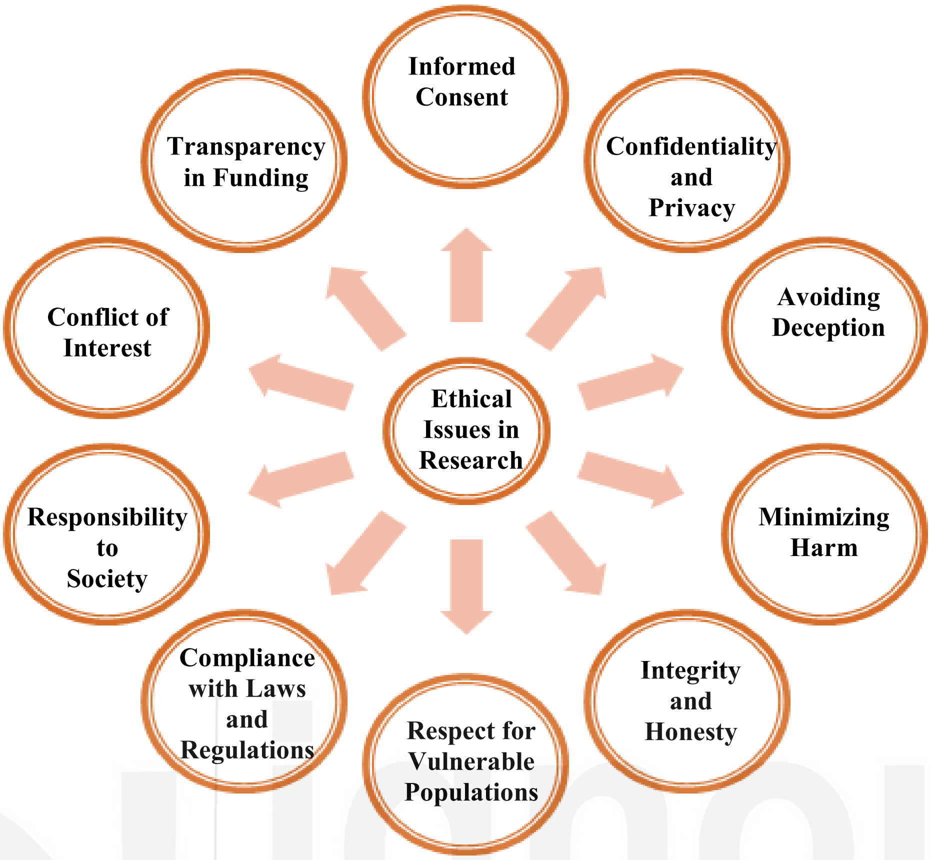


Figure 1.12: Ethical considerations in Research

Some ethical considerations business research include which are given above in the Figure 1.12 are discussed below in detail:

1. **Informed Consent:** Subjects of studies must give voluntary and informed consent before they are utilized in a study on their subject. The participant should first be informed of the general nature of the study, what will be done and what the possible downside might be.

2. **Confidentiality and Privacy:** The researcher must ensure that, participants information is private and their identities are protected. This accords with the safe custody of data and allowed sharing of data in a de-identified manner only.

3. **Avoiding Deception:** If there is need for deception it should be justified and the participants should be informed at the end of the study. Co-authors should ensure that they are very clear with study's objectives and approaches that will be undertaken.

4. **Minimising Harm:** In any research study, the researcher should ensure that the subject is not harmed physically psychologically, or emotionally. This entails, in fact, the possible risks that the research may pose to people as well as societies.

5. **Integrity and Honesty:** The researchers begin their research with a high level of professionalism in that they should not engage in fraud, deception or piracy with a view of enriching their research findings. Sticking to facts, including case data, as well as results, is crucial.

6. **Respect for Vulnerable Populations:** The latter is especially true when the subject of a study is vulnerable populations, including children or the

elderly, as well as people with low incomes. Thereby, these groups should not be exploited by researchers.

7. **Compliance with Laws and Regulations:** The subjects involved in the research and the user of the research must willingly agree to participate in the study and they should be free from influence coerced by promise of reward or fear of some penalty. However, researchers have to follow all the relevant laws, rules and regulations on use of human subjects as well as all the institutional recommendations on research ethical consideration. This include acquiring permission from ethic committees where necessary.

8. **Responsibility to Society:** People carrying out research work are supposed to be aware of the potential impact of the research, and embrace the goal of making the world a better place. This encompasses issues of utilization and dissemination of results from research activities.

9. **Conflict of Interest:** It is therefore expected of researchers to declare any interests they might have that can influence the neutrality or wisdom of the research. These refer to factors such as financial relations or affiliations that may well color or affect the results of the research.

10. **Transparency in Funding:** Scholars should declare to supporters of their studies and other likely source of bias.

Neglecting these ethical concerns might have serious consequences for the research project, thus it is important to explain how participant rights are safeguarded. We will discuss this topic in detail in unit 14.

1.10 SPECIFIED CASELETS IN BUSINESS RESEARCH

Caselets are short, concise scenarios used to illustrate business research problems or situations.

Figure 1.13: Caselets in Business Research

Here are a few examples of caselets:

1. **Market Penetration for a New Product**

   Scenario: A company, XYZ Corp, is launching a new line of eco-friendly cleaning products. They want to understand which market segments are most likely to adopt these products early. The company conducts a survey among different demographics and collects data on purchasing habits, environmental concerns, and brand loyalty.

   Research Focus:

   - Identify target demographics for the new product.
   - Assess the level of environmental concern across different age groups.
   - Determine potential barriers to adoption.

2. **Customer Satisfaction and Retention**

   Scenario: ABC Retail, a chain of clothing stores, has noticed a decline in repeat customers over the past year. They decide to investigate the root causes behind this trend. The company gathers data through customer feedback forms, mystery shopper evaluations, and analysis of sales data.

   Research Focus:

   - Measure current customer satisfaction levels.
   - Identify key factors leading to dissatisfaction or disengagement.
   - Develop strategies to improve customer retention based on findings.

3. **Effectiveness of Marketing Campaigns**

   Scenario: DEF Tech has run several digital marketing campaigns over the past six months to promote its latest software product. They want to evaluate the effectiveness of these campaigns in terms of generating leads and converting them into sales. Data is collected from web analytics, lead tracking, and sales records.

   Research Focus:

   - Analyse the ROI of different marketing channels.
   - Determine which campaign elements (e.g., email, social media) were most effective.
   - Suggest improvements for future campaigns based on the data.

4. **Product Development Feedback**

   Scenario: GHI Foods is developing a new flavour of snack chips and wants to gauge consumer interest before the official launch. They organise taste tests and collect feedback from participants on flavour preferences, packaging, and pricing.

   Research Focus:

   - Evaluate overall acceptance of the new flavour.
   - Identify any improvements needed in flavour or packaging.
   - Understand pricing sensitivity and its impact on purchase intention.

5. **Competitive Analysis for Market Entry**

   Scenario: JKL Electronics is considering entering the wearable technology market. To make an informed decision, the company conducts a competitive analysis, including reviewing existing competitors' product features, pricing strategies, and market share.

   Research Focus:

   - Map out the competitive landscape.
   - Identify strengths and weaknesses of key competitors.
   - Determine potential gaps in the market that JKL Electronics could exploit.

6. **Employee Satisfaction and Productivity**

   Scenario: MNO Services has noticed a recent dip in employee productivity and wants to understand if there's a correlation with employee satisfaction. The company conducts an internal survey to assess job satisfaction, work environment, and managerial effectiveness.

   Research Focus:

   - Identify key drivers of employee satisfaction.
   - Correlate satisfaction levels with productivity metrics.
   - Propose initiatives to improve both satisfaction and productivity.

Each caselet provides a snapshot of a business problem and outlines the research focus areas to help guide investigations and decision-making.

Check Your Progress B

1) What are the key stages of the business research process?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

2) Define the role of transparency and honesty in reporting research findings?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

3) How can case studies or caselets help in understanding the application of business research?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

4) Fill in the blanks

a) Research that starts with a general theory and moves toward specific observations is called ………………….. research.

b) A key ethical issue in business research is ensuring ………… of participants' personal data.

c) The two broad categories of research objectives are ………. and ……………. level objectives.

d) Business research plays a crucial role in improving ……………. and reducing risks in decision-making.

1.11 LET US SUM UP

Research can be defined as a diligent and systematic inquiry or investigation into a subject in order to discover or revise facts, theories, applications, etc. with regard to two types of research methods, research can be divided into inductive and deductive research. Inductive research moves from specific observations to general conclusions, while deductive research moves from general theories to specific hypotheses.

The targets of research can be general or specific referred to as micro and macro. As to the micro-objectives, one examines concrete problems or persons. On the other hand, when discussing macro-objectives, one must simultaneously examine general social or economic problems. Freely researched motivations include curiosity, the need to address well-known problems, and the opportunity for change and improvement that research presents.

Business research can also be broken down into longitudinal and cross-section research based on the length of time it looks at, exploratory, descriptive, and causal research based on its goals, and quantitative and qualitative research based on the types of data it collects. Market conditions analysis, evaluation of business propositions, and decision-making processes are central tasks of business research. It has a sinister mandate of participating in managing change and performance in organizations at all levels.

Business research involves problem definition, research questions, method determination, data collection, analysis and the conclusion. Every research work should adhere to these ethical standards to be in a position to justify the finding, validity, and informative worth. This includes matters related to first, regarding human participants, second, regarding privacy, and lastly, regarding matters of conflict of interest. Specific caselets display instances and utilization. of business research in practice where such researching was applied to solve a certain problem or answer a certain research question. The outcome of a well conducted research should always be relevant, accurate, and reliable, should show clear aims and objectives, proper identification of the method of research and clear demonstration of research findings.

1.12 KEYWORDS

Caselet: A caselet is a shorter version of a case study, usually two to three pages long. Caselets are often used as teaching tools in executive education programs and B-schools. They can describe a sequence of events or present a problem that requires decision making.

Deductive Research: Starts with a theory, develops hypotheses, and then collects and analyzes data to test those hypotheses. Deductive reasoning is narrower and more concerned with testing or confirming hypotheses. Deductive conclusions are reliable if the premises they're based on are true.

Ethics: Ethics is a system of moral principles that includes ideas about right and wrong, and how people should (or should not) behave in general and specific cases.

Inductive research: Starts with observations, looks for patterns, and then develops theories. Inductive reasoning is more open-ended and exploratory, and conclusions can't be fully proven.

Research: Research is the careful consideration of study regarding a particular concern or research problem using scientific methods.

Research Ethics: Research ethics are the moral principles that guide researchers in their work to ensure the safety and welfare of participants and the public. They also help establish the validity of research and build public trust.

1.13 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress A

1. True/False

a False

b False

c True

d True

e False

Check Your Progress B

2. Fill in the blanks

a Deductive

b Confidentiality

c Micro, Macro

d Efficiency

1.14 TERMINAL QUESTIONS

1) What is research, and why is it considered essential in academics and practical applications?
2) Discuss the role of micro and macro level of objectives in research. How do they influence study design?
3) What are the primary classifications of research? Explain with the help of examples.
4) In what ways does business research facilitate planning and strategy formulation?
5) Why is confidentiality important in the research process?
6) Discuss how ethical practices can influence the outcome of a case study in business research.

1.15 FURTHER READINGS

Bell, E., Bryman, A., & Harley, B. (2022). *Business research methods*. Oxford university press.

Hair Jr, J., Page, M., & Brunsveld, N. (2019). *Essentials of business research methods*. Routledge.

Malhotra, N. K., Nunan, D., & Birks, D. F. (2020). *Marketing research*. Pearson UK.

Note	These questions are helpful to understand this unit. Do efforts for writing the answer of these questions but do not send your answer to university. It is only for your practice.

UNIT 2 REVIEW OF LITERATURE

Structure

2.0 OBJECTIVES

After studying this unit, you will be able to:

- Discuss the purposes of review of literature;
- Explain the sources of review of literature;
- Identify different types of literature; and
- Learn about the review of the literature writing process.

2.1 INTRODUCTION

In the previous unit, we explored the business research process, which provided an understanding of how to systematically approach and structure research. Building on that foundation, this unit focuses on the Review of Literature (ROL), a crucial step in the process. ROL helps in identifying research gaps, refining objectives, and ensuring the study is connected in existing knowledge, thereby laying a strong groundwork for effective research.

In the world of research, a ROL serves as the foundation upon which every great study is built. Think of it as a map that guides researchers through the vast landscape of existing knowledge, helping them find their way towards new discoveries and insights. By carefully examining what has already been written on a topic, researchers can identify gaps in knowledge, refine their ideas, and build a stronger argument for their own work.

For learners stepping into the field of business research, understanding how to conduct a thorough review of literature is a crucial skill. It's more than just summarising past studies; it's about connecting ideas, contrasting perspectives, and finding patterns that can inspire fresh ways of thinking.

This unit will guide you through the purpose of a literature review, the essential strategies for conducting it, types of literature review and the major sources of information that you can use to create a comprehensive understanding of your research topic.

As you navigate through this unit, you'll discover practical tools and techniques that will not only help you build a solid research foundation but also sharpen your analytical skills. Whether you are preparing a research proposal, writing a dissertation, or simply exploring a topic of interest, mastering the art of reviewing literature will empower you to engage with research in a more meaningful and insightful way.

Let's embark on this journey to understand the importance of literature review in research, and how it can shape the path to new knowledge and innovations.

2.2 PURPOSE OF ROL

A literature review is part of the research process. It provides considerable information on the topic being researched and the various works that had gone on in the field over the years. These materials are gathered by the researcher from many sources such as journals, books, documents etc. The review of such a literature could be a matter-of-fact presentation of the information or it could be a synthesis of a large amount of information and put together subject wise for the purpose of understanding. It can be just a simple summary of the sources, but it usually has an organisational pattern and combines both summary and synthesis. In summary all the information is synthesised and given in a capsule form. It synthesises and organises the entire information in terms of its relevance and appropriateness to the topic of research. It might give a new interpretation of old material or combine new with old interpretations. Or it might trace the intellectual progression of the field, including major debates. And depending on the situation, the literature review may evaluate the sources and advise the reader on the most pertinent and relevant information. Figure 2.1 gives a broad overview of Purpose of ROL.

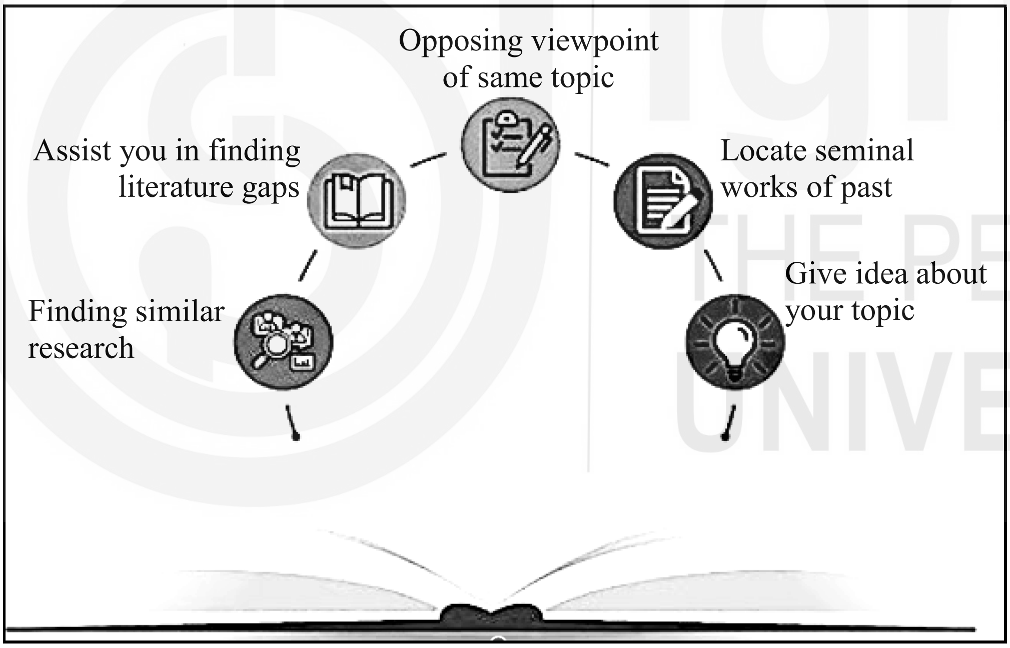


Figure 2.1: Purpose of ROL

Difference between Literature review and Academic research paper

The question arises as to how the literature review differs from an academic research paper. While the main focus of an academic research paper is to develop a new argument, a research paper will contain the literature review as one of its chapters. In a research paper one uses the literature as a basic foundation and support for newer ideas and insights into the research topic of interest. Literature review on the other hand summarises and synthesises the many arguments and literature and research findings gathered from such a review and puts forward arguments in favour or against the particular topic and its findings.

Materials to be included in review of literature

The next question is regarding how many and how much of materials to be included in review of literature. There is no hard and fast rule about this. The researcher has to definitely include the materials from classic and pioneering works in the area. In addition, the researcher should also include all the relevant research works published more recently especially in the last 5 to 10 years.

As for the types of sources to be consulted for review of literature, this includes books, journal articles, monographs, documents, grey literature such as unpublished documents or research papers read at some conferences etc. All of these sources you will be studying later in this unit. In addition, the internet is an important source from which articles and abstracts could be downloaded for this purpose.

Once all the materials have been gathered from different sources as mentioned above, the researcher should organise the same according to the year of publication and the subject matter must be organised to give meaning to the entire literature gathered keeping in view the present research topic of interest to the researcher.

The researcher can evaluate these materials on the basis of the methodology used, the research findings arrived etc. The researcher could also include in such a review certain minimal and directly relevant historical accounts regarding the research topic.

The specific purposes of a Review of the Literature are given below:

a) Identifying variables relevant for research

When the researcher makes a careful Review of the Literature, he becomes aware of the important and unimportant variables in the concerned area of research. A careful Review also helps the researcher in selecting the variables (refer below notes for more clarity) lying within the scope of his interest, in defining and operationalising as well as in identifying variables which are conceptually and practically important. Thus, a Review of the Literature, on the whole, prepares the researcher to formulate a research problem in which conceptualisation and practically important variables are selected.

b) Avoidance of repetition

A Review of the Literature helps the researcher in avoiding any duplication of work done earlier. A careful review always aims at interpreting prior studies and indicating their usefulness for the study to be undertaken. Thus prior studies serve as the foundation for present research. In some cases the duplication or replication of prior studies becomes essential. This is especially true when the researcher wants to test the validity of the earlier studies. In such a situation too, a careful review helps the researcher in getting acquainted with the number and nature of the studies related to the present research whose validity is being assessed at present.

c) Synthesis of prior works

Review of the Literature enables the researcher to collect and synthesise prior studies related to the present study. This, in turn, helps the researcher in building a better perspective for future research. A synthesised collection of prior studies also helps a researcher to identify the significant overlaps and the gaps among the prior works.

d) Determining meaning and relationship among variables

A careful Review of the Literature enables the researcher in discovering important variables relevant to the area of the present research. When significant variables are discovered, the relationship among them can be identified. Subsequently, the identified relationship is incorporated into different hypotheses. Thus, for conducting a scientific study, the relationship between the different variables must be explored by reviewing the literature so that a good context may be built up for subsequent investigations.

In addition to these specific purposes, there are some general purposes of the literature review also. These are:

- To argue for the relevance and the significance of the research question.
- To provide the context for one's own methodological approach
- To establish one's own credibility as a knowledgeable and capable researcher.
- To argue for the relevance and appropriateness of one's own approach.

For more clarity

Variables are characteristics or properties that can change or vary. They are what researchers' study measure or manipulate to understand relationships, patterns, or effects. It can be:

Independent variable: The variable that is changed or controlled by the researcher to see its effect on other variables. For example: If you're studying the impact of study hours on exam performance, the study hours is the independent variable.

Dependent Variable: The variable that is measured to see how it is affected by the independent variable. For example: In the above example, exam performance is the dependent variable because it depends on the study hours.

2.3 IDENTIFICATION OF RESEARCH AREA

The process of identifying a research area is closely connected to the review of literature, as reviewing existing studies helps in discovering potential topics, understanding current trends, and identifying research gaps. By exploring what has already been studied, you can refine your focus and ensure your research contributes meaningfully to the field. Identifying a research area means figuring out what topic you want to study or learn more about. This step is like choosing a focus for your research, so you don't get

lost in too much information. To find the right research area, you can follow these simple steps:

1. **Think About Your Interests:** Start by thinking about what topics or problems interest you the most. It could be something you're curious about, like a trend in technology, a social issue, or a problem in your industry.

2. **Do Some Background Reading:** Read a little about different topics to see which one has enough information available. Look for articles, books, or reports to get an idea of what other people have already studied in those areas.

3. **Check for Gaps:** Try to find out what hasn't been studied yet. Look for questions that haven't been fully answered or areas where more information is needed. This helps you find a topic that's new or hasn't been explored much.

4. **Relate it to Real-Life Problems:** Think about how the topic connects to real-world issues or challenges. Research is most useful when it helps solve a problem or makes life easier for people or businesses.

5. **Choose a Specific Focus:** It's better to pick a topic that's not too broad. For example, instead of just studying "business," you could focus on "how small businesses use social media for marketing." This makes your research clearer and more manageable.

6. **Consider Feasibility:** Think about whether you can realistically study the topic you choose. Consider factors like time, resources, access to data, and your research skills. For example, studying "consumer behaviour in rural areas" may require travel and data collection, so ensure its practical for you.

7. **Explore Emerging Trends:** Look into recent trends, innovations, or advancements in your field of interest. Emerging topics are often under-researched and can provide unique opportunities for meaningful contributions. For example, studying the impact of Artificial Intelligence on consumer decision-making is highly relevant today.

8. **Check Theoretical and Practical Relevance:** Ensure your topic is both theoretically significant (contributing to academic knowledge) and practically useful (addressing real-world problems). For example, studying "employee motivation" might use theoretical models while providing actionable insights for HR managers.

9. **Reflect on Ethical Considerations:** Ensure your research topic adheres to ethical guidelines. Avoid sensitive or controversial areas unless you are confident in addressing ethical concerns, like privacy issues in data collection or participant consent.

10. **Ask for Feedback:** If you're still unsure, talk to your teachers, friends, or experts in the field. They can help you choose a topic that's interesting and worth studying.

11. **Validate With Literature Reviews:** Start reviewing literature early to ensure there's enough prior research to build upon. Identify key studies, authors, and journals in your area to validate the relevance and importance of your chosen topic.

2.4 4 Cs OF LITERATURE REVIEW

Once you've identified your research area, the next step is to dive into the existing studies related to your topic. To do this effectively, you can use the Four Cs of Literature Review: **Cite, Contrast, Compare, and Connect**. These 4Cs of literature review provide a systematic way to synthesize existing research. These steps not only organise your findings, analyse what other researchers have said, and make sense of how their work relates to your study but also ensure your review critically evaluates the literature and positions your study effectively within the field.

Cite	Contrast
Compare	Connect

Figure 2.2: Four Cs of Literature Review

2.4.1 Cite

Citing means acknowledging the sources of information you include in your research.

By citing, you give credit to the original authors and demonstrate that your study is built on a solid foundation of existing knowledge. It also helps to show that you are not just expressing opinions but are relying on verified research.

Whenever you mention ideas, facts, or data from other studies, include the authors' name, the publication year, and other relevant details to let readers know where the information comes from.

You can use citation tools like Zotero or EndNote to give citations and keep track of references.

For instance, if you are studying the impact of social media on mental health, you might cite studies like "Smith (2022), which found that increased social media use correlates with anxiety."

2.4.2 Contrast

Contrasting involves identifying the differences between the findings, methods, or perspectives of various research studies. Pointing out these differences helps you see where researchers disagree or have found varying results. It allows you to explore the reasons behind these differences and understand the diverse views on your topic. Look closely at how different

studies approach similar problems and note where their conclusions or methods diverge.

One can create a comparison table to clearly identify differences in methods or results.

For example: "While Smith (2025) found a positive correlation between social media use and anxiety, Brown (2024) argued that social media has no significant impact on mental health in teenagers."

2.4.3 Compare

Comparing is the process of finding similarities between different research papers. When you compare studies, you highlight common patterns, trends, or agreements among researchers. This step helps you identify the most consistent findings and understand the dominant theories in your area of research. Group studies with similar results or approaches and explain how these similarities contribute to a deeper understanding of your research topic.

For example: “Both Smith (2025) and Davis (2024) found that screen time significantly affects sleep quality, indicating consistent patterns in their findings."

2.4.4 Connect

Connecting is about linking the information from existing studies to your own research. This step helps you explain how your study fits into the broader field of knowledge and how it builds on or challenges what is already known. It also shows where your research can fill gaps or answer unresolved questions.

Clearly show how the studies you reviewed relate to your research question and how they support or shape your argument or hypothesis.

For example: "Building on the work of Smith (2022), this study explores how specific social media platforms like Instagram uniquely impact anxiety levels in adolescents."

By applying the 4C’s—Cite, Contrast, Compare, and Connect—one can ensure their literature review is not just a summary of studies but a critical synthesis that positions they research within the broader academic conversation.

2.5 MAJOR SOURCES OF ROL

To conduct a thorough literature review, it's essential to gather information from reliable and varied sources. These sources provide the foundation for understanding what has already been researched and where gaps might exist in your chosen field. By exploring different types of sources, you can gain a comprehensive view of your topic and ensure that your research is grounded in well-established knowledge. The major sources of literature review include books, reports, dissertations, research databases, and journals. Each of these sources contributes valuable insights that help you build a solid framework

for your study and connect your research to the broader academic conversation.

2.5.1 Books

Books are fundamental sources for a literature review, providing in-depth insights into specific topics or broader research areas. They often serve as a foundation for theoretical understanding and conceptual frameworks. Books are especially useful for understanding historical perspectives, theoretical concepts, or foundational studies in your field. Unlike journals, books might not always have the latest research but are valuable for building the background knowledge necessary for your study.

2.5.2 Reports

Reports are a significant source of information in the review of literature. They provide detailed findings, data analysis, and insights from various organisations, institutions, and government bodies. Reports can be annual, research-based, technical, or industry-specific, and they often present real-world data and trends that support evidence-based conclusions. Utilising reports in research helps to incorporate credible and up-to-date information, offering a practical perspective on the topic under study.

2.5.3 Dissertation and Repositories

Doctoral dissertations and theses repositories are excellent sources for literature reviews, often containing in-depth analyses and detailed reviews of existing research. University libraries and online repositories house numerous dissertations that can be accessed for relevant information. These documents typically include key chapters like an introduction, literature review, research methodology, results, and discussion, providing valuable insights into prior studies. Online resources like Dissertation Abstracts International or country-specific surveys (e.g., Shodhganga in India for thesis) help researchers locate and access such dissertations easily.

Figure 2.3: Shodhganga

Shodhganga is a digital repository of Indian theses and dissertations maintained by the **INFLIBNET Centre**. It provides researchers with free access to a vast collection of doctoral work from Indian universities, offering valuable insights and a comprehensive review of existing research in various fields. This platform is an excellent resource for identifying research gaps and building a strong foundation for your study.

2.5.4 Databases

Databases act as comprehensive online repositories for scholarly articles, journals, books, and other academic resources. Platforms such as **Google Scholar, JSTOR, Scopus, and PubMed** enable researchers to access reliable and high-quality information relevant to their topics. These databases are essential for finding peer-reviewed studies, systematic reviews, and the latest research, ensuring a thorough and efficient literature review process.

2.5.5 Journal

Journals are one of the most critical and widely used sources for conducting a literature review. Academic journals contain peer-reviewed articles, which means that the content is rigorously evaluated by experts in the field before publication, ensuring credibility and high-quality information. Journals typically present original research, critical reviews, theoretical discussions, or case studies, offering insights into the latest advancements, methodologies, and findings in a specific domain.

One of the key advantages of journals is their timeliness; they provide up-to-date information, unlike books, which may take years to publish. Journals often focus on specialized topics, making them an indispensable resource for exploring niche areas of research. For instance, journals like *Nature*, *The Lancet*, *Harvard Business Review*, and *Journal of Marketing Research* are well-known for their contributions to management, healthcare, business, and marketing, respectively.

Accessing journals through reliable platforms such as **Springer**, **Elsevier**, **Wiley, Taylor & Francis**, or open-access sources like **DOAJ (Directory of Open Access Journals)** can help researchers find highly relevant studies. By reviewing journal articles, researchers can identify trends, recurring themes, and consistent findings that are essential for understanding the current state of knowledge in their field. Additionally, journals often include a detailed methodology section, which can serve as a guide for designing one's own research.

Check Your Progress A

2. What are the four Cs of Literature Review?

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

2) How would you identify the research area?

…………………………………………………………………………

…………………………………………………………………………

…………………………………………………………………………

…………………………………………………………………………

3) What is the purpose of reviewing literature?

…………………………………………………………………………

…………………………………………………………………………

…………………………………………………………………………

…………………………………………………………………………

4) What are major sources of literature review?

…………………………………………………………………………

…………………………………………………………………………

…………………………………………………………………………

…………………………………………………………………………

…………………………………………………………………………

5. Mark the correct answers:

i) What is the primary purpose of a Review of Literature (ROL)?

A. To summarize a single study

B. To identify gaps and provide a foundation for research

C. To write a conclusion for the study

D. To replace data collection

ii) A well-conducted Review of Literature does which of the following?

A. Synthesizes and organizes existing information

B. Proposes new theories without reviewing prior work

C. Focuses only on historical studies

D. Avoids evaluating the relevance of sources

iii) Which of the following is considered grey literature?

A. Peer-reviewed journal articles

B. Conference papers and government reports

C. Books published by academic presses

D. Articles indexed in Scopus

iv) What does "Contrast" in the 4Cs of literature review mean?

A. Highlighting differences in research findings or methods

B. Identifying recurring themes in studies

C. Citing multiple sources on the same topic

D. Connecting the literature to personal opinions

2.6 CLASSIFICATION AND INDEXING OF JOURNAL

Journals are among the most essential and often used resources for doing a literature review. Journals are often categorised according to their discipline focus (e.g., medical, business, or multidisciplinary), the kind of research published (e.g., empirical paper, Literature reviews paper, or theoretical analyses), and their influence and prestige (e.g., high-impact factor journals such as Nature). This categorisation assists researchers in identifying articles that are most relevant to their area and objectives.

Journals enhance their legitimacy and exposure by indexing, which involves inclusion in specialist databases such as Scopus, Web of Science, PubMed, or DOAJ. Indexed journals adhere to rigorous quality requirements and are more often subjected to peer review, guaranteeing superior information quality.

Scopus is a multidisciplinary database that catalogues hundreds of high-quality articles, while PubMed specializes in life sciences and biomedical research.

The quality and effect of journals are often assessed by metrics such as Impact Factor (IF), which quantifies the average citations received by a journal's articles, and h-index, which evaluates both productivity and citation impact. High-impact journals are often more esteemed and frequently referenced, making them significant resources for a literature study.

Furthermore, journals may be classified as open-access or subscription-based. Open-access publications, such as *sustainability*, provide unrestricted access to their material, on the other hand subscription-based journals, like *Journal of Consumer behaviour*, often possess better impact factors and established reputations.

By comprehending the categorisation and indexing of journals, researchers may identify the most reputable, relevant, and influential sources to provide a robust basis for their literature study.

For more clarity

Journals can be classified into different categories, such as:

Peer-Reviewed Journals: These journals have a rigorous review process where experts evaluate submitted articles for quality and validity before publication. They are considered highly credible and reliable sources.

Open Access Journals: These journals provide free access to their articles, making research available to a broader audience. While many open-access journals are reputable, some may lack rigorous peer-review processes.

Specialised Journals: These focus on specific areas of study within a broader discipline. For example, a journal may concentrate on finance, marketing, or technology, allowing researchers to find niche topics more easily.

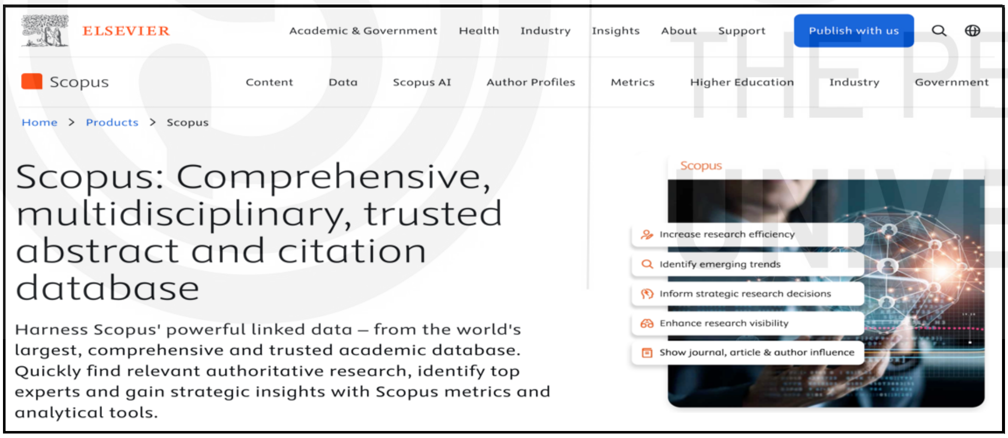
Indexing refers to the process of including academic journals in recognised databases that catalogue and provide access to journal articles. Indexed journals are considered more credible and widely recognized because they meet specific quality and peer-review standards. These databases make it easier for researchers to discover, access, and cite scholarly work, thereby enhancing the visibility and impact of research publications. Some of the most prominent indexing systems include:

2.6.1 Scopus

Scopus is one of the largest and most comprehensive abstract and citation databases, covering a wide range of disciplines, including science, technology, medicine, social sciences, and arts and humanities. It is known for its rigorous selection criteria that ensure only high-quality and peer-reviewed journals are included. Scopus also provides tools for tracking citation metrics, h-index scores, and trends in academic publishing, making it a valuable resource for researchers, academicians, and institutions worldwide.

Its strengths:

- It covers over 25,000 titles from more than 5,000 publishers globally till 2024.
- It provides metrics like CiteScore and SNIP (Source Normalized Impact per Paper) to evaluate journals.
- It enables researchers to identify the most impactful studies in their field.



Source: Scopus **Figure 2.4: Scopus**

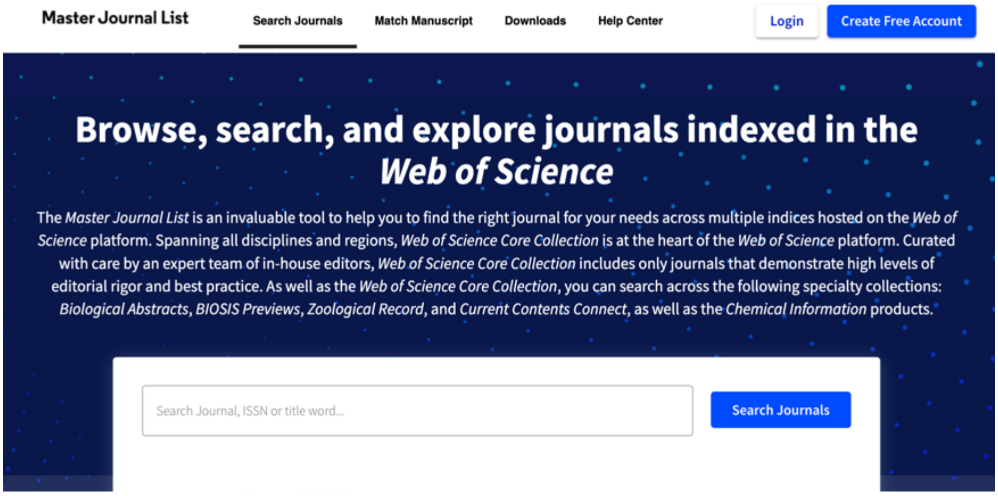
2.6.3 Web of Science

The Web of Science (WoS), formerly referred to as Web of Knowledge, is a leading indexing service. It provides access to a collection of citation databases including academic journals, conference proceedings, and other intellectual papers across several fields. WoS emphasizes high-impact journals and provides tools like citation monitoring, enabling researchers to comprehend the interconnections among studies and identify the most significant publications.

Its strengths:

- Access to Journal Citation Reports (JCR), which offers Impact Factor statistics.

- Includes disciplines from engineering and natural sciences to the arts and humanities.
- Facilitates interdisciplinary study by connecting citations across many fields.



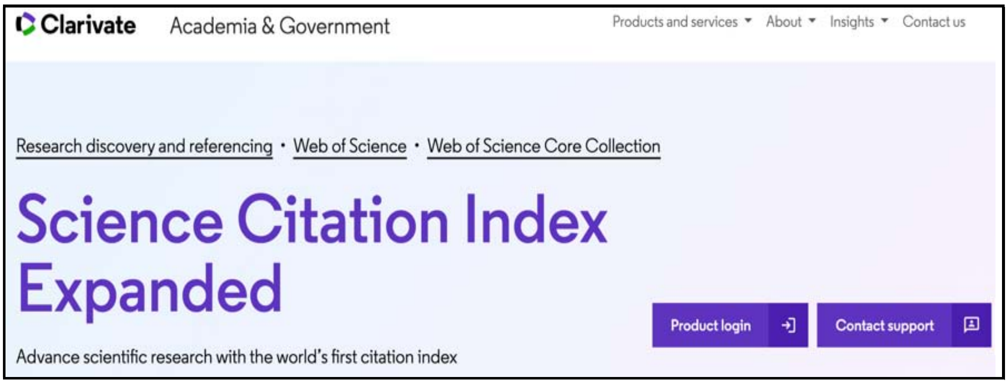
Source: Web of science **Figure 2.5: Web of Science**

2.6.4 SCI

The Science Citation Index (SCI), now expanded into the Science Citation Index Expanded (SCIE), is among the most esteemed indexing systems for scientific publications. It is included in the Web of Science and managed by Clarivate Analytics. The SCI emphasizes publications in scientific and technological fields, guaranteeing high-quality and influential research.

Its strength:

- Monitors article citations and recognizes significant research works,
- Comprises hundreds of periodicals in the fields of natural sciences, technology, and medicine.
- Frequently used as a standard for assessing the quality of publications in scientific fields.



Source: Clarivate **Figure 2.6: Science Citation Index Expanded**

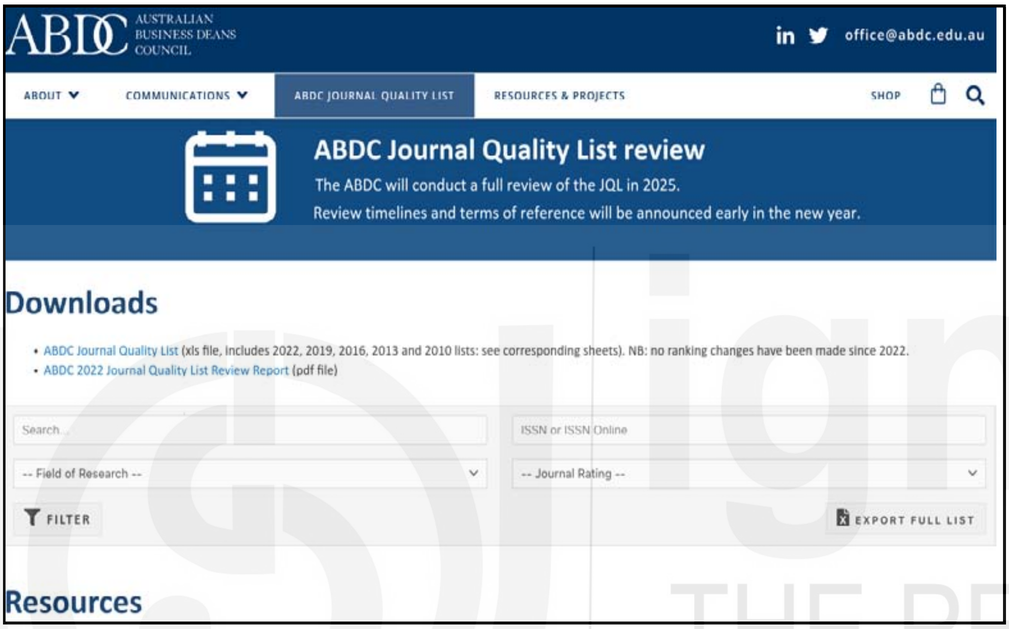
2.6.5 Australian Business Deans Council (ABDC)

The Australian Business Deans Council (ABDC) offers a categorisation system specifically for journals in business, management, and economics. The ABDC method classifies journals into four quality tiers: A*, A, B, and

C, with A* being the highest category. It serves as a crucial resource for researchers in business-related fields to pinpoint high-impact publications.

Its Strengths:

- Concentrated on business and management domains, providing a specialist viewpoint.
- Aids researchers in identifying articles that conform to academic and industrial criteria.
- Encourages publication in esteemed publications to preserve academic integrity.



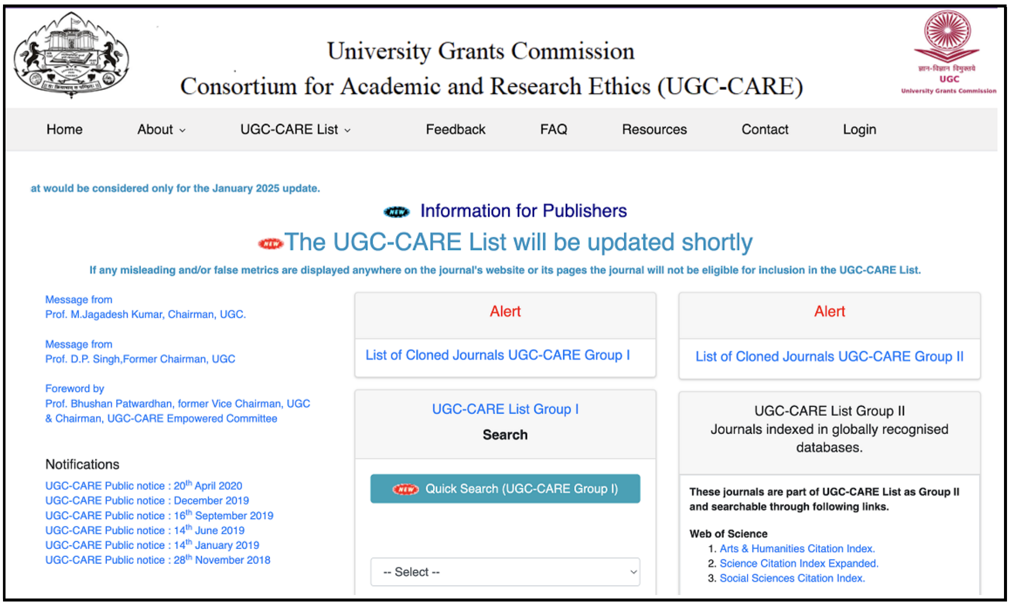
Source: ABDC **Figure 2.7: ABDC Journal list**

2.6.6 UGC CARE List

The UGC CARE (Consortium for Academic and Research Ethics) is an initiative by the University Grants Commission (UGC) in India to maintain high-quality research standards. It provides a list of trusted and reputable journals across different fields, ensuring that researchers and students publish their work in genuine and credible sources. The UGC CARE list is regularly updated to include journals that follow ethical publishing practices, promoting integrity and quality in academic research.

Its Advantages:

- Helps researchers, especially in India, identify legitimate and high-quality journals for publication.
- Regularly updated to include journals that meet ethical and academic criteria.
- Prevents predatory publishing by maintaining a vetted list of reputable sources.



Source: UGC-Care **Figure 2.8: UGC Care List**

2.7 PROCESS OF REVIEW OF LITERATURE

Conducting a literature review is a systematic process that involves identifying, analysing, and synthesising relevant sources to understand the current state of knowledge in your research field. A well-structured review not only summarizes existing studies but also critically evaluates them to identify themes, debates, and gaps, laying a solid foundation for your research.

Before starting literature review, it is essential to understand the different types of reviews to align them with your research goals.

Types of Review of Literature

A *Narrative Literature Review* provides a comprehensive summary of existing literature without adhering to a specific methodology, offering insights into key concepts and developments in a field.

In contrast, a *Systematic Literature Review* uses a structured and transparent methodology for selecting and analysing studies, ensuring a balanced and reliable synthesis of knowledge.

A *Meta-Analysis* takes this further by employing statistical methods to combine and analyse quantitative results from multiple studies, providing a statistical synthesis of outcomes. For broader research areas, a *Scoping Review* maps existing literature without strict inclusion criteria, identifying gaps and shaping the direction for future studies.

A *Critical Literature Review*, on the other hand, evaluates the strengths and weaknesses of existing research, highlighting contradictions, methodological concerns, and areas for further investigation.

Finally, an *Integrative Literature Review* synthesizes diverse sources, including theoretical and empirical studies, to construct a unified understanding of the topic, bridging gaps between past and present research

to explore future directions. Understanding these types ensures the review aligns with the intended purpose and research objectives. Once you understand the type of literature review, you need to understand the process of Review of Literature. Below is a step-by-step guide to writing an effective literature review (Refer Figure 2.9)

Figure 2.9: Literature Review in 7 steps

Step 1: Define the Scope

The first step in writing a literature review is to define its scope clearly, ensuring it remains focused and manageable. This involves determining exactly what you want to explore and setting boundaries for your review.

- **Clarify Your Research Question:** Narrow down your research focus by asking, "What specifically am I investigating?"
- **Set Boundaries:** Decide on the time frame of studies to include (e.g., recent vs. historical) and the types of sources to consider (e.g., peer-reviewed journals, books, or reports). For instance, if researching the impact of social media on teenagers' mental health, you might limit your review to peer-reviewed articles from the last five years while excluding non-academic sources like magazines.

Step 2: Search for Relevant Literature

Once your scope is defined, the next step is to search for relevant studies and materials.

- **Utilize Academic Databases:** Search platforms such as PubMed, JSTOR, or Google Scholar using specific keywords related to your topic.
- **Check Annotated Bibliographies:** Explore references in key articles to find additional relevant sources.
- **Use Search Filters:** Refine your searches using filters (e.g., publication date, study type) to ensure the relevance of results. This systematic search process helps in gathering a wide range of reliable sources for your review.

Step 3: Evaluate Sources

Critically assess the quality and reliability of the sources you've collected to ensure their credibility.

- **Author Credibility:** Verify the expertise of the author, such as their qualifications or prior work in the field.
- **Publication Source:** Prefer peer-reviewed journals and reputed publishers, avoiding non-academic sources.
- **Publication Date:** Ensure the studies are recent enough to reflect current knowledge unless historical context is necessary. For example, a study on teenage mental health authored by a clinical psychologist and published in a peer-reviewed journal is more credible than a personal blog.

Step 4: Organize and Summarize

Organize your findings systematically to uncover patterns and relationships.

- **Group Sources by Themes:** Categorize studies based on shared topics or ideas to better understand the body of literature.
- **Summarize Key Points:** Write concise summaries highlighting the main findings, methodologies, and relevance of each source.
- **Identify Recurring Ideas:** Look for consistent patterns or differing viewpoints among studies. This step ensures your literature review builds a cohesive narrative and highlights critical insights.

Step 5: Identify Themes, Debates, and Gaps

Dig deeper into your organized sources to identify the main ideas, ongoing debates, and research gaps.

- **Identify Themes:** Look for recurring ideas across studies that provide a cohesive framework for your review.
- **Spot Debates:** Highlight areas where researchers disagree or present conflicting results.
- **Find Gaps:** Identify underexplored aspects or unanswered questions that can justify your research. For example, while many studies may agree on social media's role in anxiety, there may be debates about its influence on depression, and a lack of studies on specific features like filters may present a gap.

Step 6: Create an Outline

Organize your insights into a structured outline that guides your writing process.

- **Overview:** Start with a brief introduction outlining the research problem and purpose of the review.
- **Organize by Themes:** Structure your main sections around the themes and debates identified earlier.
- **Use Subheadings:** Break down each section for clarity, such as "Theme 1: Social Media and Anxiety" or "Debate: Social Media and Depression."

- **Consider Order:** Choose a logical flow, whether chronological, methodological, or thematic (see below more clarity). A well-structured outline makes the writing process smoother and ensures your review is coherent.

Step 7: Write Your Literature Review

Finally, write the literature review following your outline and incorporating critical analysis.

1. Introduction
   - Introduce the research topic and explain its relevance.
   - Highlight any gaps or controversies in the literature.
   - Provide a preview of the themes or questions to be explored.
2. Body
   - **Follow Your Outline:** Develop each section based on the themes or aspects identified earlier.
   - **Summarize and Analyse:** Combine summaries of findings with your own critical evaluation.
   - **Connect Ideas:** Link studies to show relationships and build a cohesive narrative.
3. Critical Analysis
   - Evaluate methodologies and highlight the strengths and weaknesses of studies.
   - Address potential biases, such as cultural or publication biases, in the literature.
   - Emphasize gaps and debates to justify your research.
4. Conclusion
   - Summarize the main themes and arguments discussed.
   - Reiterate the significance of your research and its contribution to the field.
   - Briefly connect the review to the next steps in your research, such as methodology or specific hypotheses.

Thus, writing the literature review is a complex task which can be made easy by following the above-mentioned steps meticulously.

For More Clarity

1. Chronical approach of doing literature review: The chronological method organizes the literature based on the timeline of research. In this approach start with the latest studies in the field.
2. Methodological approach of doing literature review: The methodological method organizes the literature based on the methods used in the studies. It is particularly useful when evaluating studies with varying research designs like qualitative and quantitative research design.

3. Thematic approach of doing literature review: The thematic method organizes the literature based on themes, concepts, or key issues in the research. This approach is effective when multiple themes or categories are being explored. For example Topic: Themes in Leadership Styles and Their Impact on Organizational Success include theme like Transformational Leadership, Transactional Leadership, and Servant leadership.

Check Your Progress-B

1) Discuss the strategies in writing a review.

....

....

....

....

....

2) What are the steps to be followed in the writing process of a review of literature?

....

....

....

....

....

3) Why is feedback important in review of literature?

....

....

....

....

....

4) Is redrafting necessary in review of literature? What advantage does it have?

....

....

....

....

2.8 SAMPLE TEMPLATE

The sample template is a structured framework designed to systematically organize and summarize the findings of the literature review. This structure ensures consistency, thorough documentation, and ease of comparison across studies. Below is the table format followed by detailed explanations for each column.

Table 2.1: Sample Template of Literature Review

Sl. No.	Year	Authors	Title	Research Methodology	Techniques of Data Analysis	Main Findings	Identification of Research Gaps
1	**2024**	John Doe, Jane Smith	The impact of technology on Marketing	Quantitative	Regression Analysis	Technology significantly increases marketing efficiency	Need for more studies on the impact of emerging technologies like AI on marketing strategies.
2	**2023**	Alice Johnson, Tom Lee	Consumer Behavior in E-commerce	Qualitative	Thematic Analysis	Consumers value convenience over price when shopping online.	Limited research on the role of personalization in consumer decision-making
3	**2022**	Rahul Gupta, Maria Wong	Financial Literacy and Investment Habits	Mixed Methods	Statistical Analysis & Interviews	Financial literacy positively affects investment decisions.	Lack of studies focused on financial literacy among different age groups and demographics.

2.8.1 Year

The year in which the study was published, helping to understand the timeline of research development. For example: Studies from 2020 onward may address the effects of the COVID-19 pandemic, while earlier studies might focus on pre-digital or pre-globalized trends. The important of year is that it helps in tracking the timeline of studies helps identify shifts in trends, and priorities over time.

2.8.2 Author(s)

Name the Author who conducted the study, attributing proper credit and allowing further exploration of their other works. For example: an author like "John Doe" might frequently publish on Artificial Intelligence in marketing, making their works a cornerstone for the review. Understanding key authors in the field can help to identify influential researchers and collaboration networks.

2.8.3 Title

Title summarizes the primary focus or subject of the research study in a concise format. For example: A title like "The Role of Big Data in Predicting Consumer Behaviour" gives immediate insight into the scope and objectives of the study. A clear title helps to categorize studies based on relevance, and making it easier to align them with specific research questions.

2.8.3 Research Methodology

Research methodology describes the approach or design of the study, indicating how the data was collected, analysed, and interpreted. Types of Research methodology include qualitative, quantitative and mixed method. Quantitative Research Method focuses on numerical data and statistical analysis. On the other hand Qualitative Research method focus on interview, observation and not on numerical data. Mixed Research method combines both quantitative and qualitative approaches for a holistic understanding. For example: A qualitative study may use interviews to explore consumer behaviour, while a quantitative study may use a survey and regression analysis to identify trends.

2.8.4 Techniques of Data Analysis

This includes tools or methods used for analysing the data, indicating the study's depth and analytical approach. You will further learn different data analysis technique later in this course.

2.8.5 Identification of Research Gaps

It highlights areas where further research and identifies opportunities for the new contributions. For example, “Limited exploration of AI in small-scale business marketing” highlights an underexplored area that can shape the researcher’s focus. Identifying Research gaps justifies the current research’s relevance and positions it to address unanswered questions or limitations in existing literature.

2.9 LET US SUM UP

The Review of Literature (ROL) serves as a critical foundation in the research process, enabling researchers to explore existing knowledge, identify gaps, and refine their research objectives. This unit highlights the importance of ROL as a tool to synthesize and evaluate prior studies, providing a contextual framework for new research. A literature review is not merely a summary of previous works but a comprehensive analysis that connects ideas, contrasts perspectives, and identifies patterns and inconsistencies within the field. By doing so, it helps researchers build credibility, formulate hypotheses, and establish a solid foundation for their study.

The unit begins by emphasizing the purpose of ROL, which includes synthesizing relevant studies, avoiding repetition, and determining meaningful relationships among variables. It contrasts ROL with academic research papers, noting that while the latter develops new arguments, ROL consolidates existing insights to support the research question. The materials included in ROL range from classic and pioneering works to recent studies, sourced from books, journals, reports, dissertations, and databases like Scopus and Web of Science. The unit also explains the critical role of identifying research areas by exploring topics of interest, assessing research gaps, and ensuring theoretical and practical relevance.

To structure a literature review effectively, the unit introduces the Four Cs—Cite, Contrast, Compare, and Connect—which help organize findings and establish the study's position within the broader academic conversation. It also discusses different types of ROL, including Narrative, Systematic, Meta-Analysis, Scoping, and Critical reviews, each tailored to specific research objectives. A systematic process for conducting ROL is outlined, starting from defining the scope, searching for relevant literature, and evaluating sources, to organizing findings, identifying themes, and drafting the review. The unit elaborates on approaches like chronological, thematic, and methodological structuring, providing examples to illustrate their application.

Major sources of ROL such as books, journal articles, dissertations, and databases are explored in detail, with a focus on their strengths and relevance to specific research needs. Journals are classified based on indexing systems like Scopus, Web of Science, and UGC CARE List, ensuring credibility and quality. Additionally, a sample template is provided to guide researchers in systematically documenting their findings, covering key elements like year, authors, title, research methodology, main findings, and research gaps.

In conclusion, this unit equips researchers with the tools and strategies needed to conduct an effective literature review, emphasizing the importance of critical analysis, structured synthesis, and alignment with research goals. By mastering the art of reviewing literature, researchers can build a robust foundation for their studies and contribute meaningfully to their respective fields.

2.10 KEYWORDS

Cite: Acknowledging sources used in research to credit original authors and demonstrate the basis for building new arguments.

Grey Literature: Non-peer-reviewed sources such as conference papers, reports, theses, and government documents that contribute valuable, albeit informal, insights to a literature review.

Indexing: The process of including academic journals in recognized databases (e.g., Scopus, Web of Science) to ensure credibility, accessibility, and quality assurance of published articles.

Repository: A repository is a centralised location or storage system where data, files, or resources are systematically organized, stored, and managed for easy access, retrieval, and use.

Review of Literature (ROL): A systematic examination and synthesis of existing studies on a specific topic, aimed at identifying gaps, summarizing knowledge, and establishing a foundation for new research.

Research Gap: An unexplored or insufficiently studied area within a field of research, offering opportunities for further investigation and contribution.

2.11 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress A:

5. (i) (a) (ii) (b) (iii) (c) (iv) (a)

2.12 TERMINAL QUESTIONS

1) What are the primary purposes of a Review of Literature? How does it help in avoiding duplication and identifying research gaps?
2) What are the advantages of using indexed journals in a literature review? Highlight the role of Scopus and the UGC CARE List in ensuring research quality.
3) What are the major sources of literature used in ROL? Briefly describe how books, journals, dissertations, and grey literature contribute to a comprehensive review.
4) Describe the process of identifying a research area. How does reviewing existing literature help in refining research questions and ensuring relevance?
5) Explain the 4Cs of Literature Review (Cite, Contrast, Compare, Connect) in detail.

2.13 FURTHER READINGS

1. Dattakumar, R., & Jagadeesh, R. (2003). A review of literature on benchmarking. Benchmarking: An International Journal, 10(3), 176-209.
2. Cooper, C., Booth, A., Varley-Campbell, J., Britten, N., & Garside, R. (2018). Defining the process to literature searching in systematic reviews: a literature review of guidance and supporting studies. *BMC medical research methodology*, *18*, 1-14.
3. Paul, J., Khatri, P., & Kaur Duggal, H. (2024). Frameworks for developing impactful systematic literature reviews and theory building: What, Why and How?. *Journal of Decision Systems*, *33*(4), 537-550.

Note	These questions are helpful to understand this unit. Do efforts for writing the answer of these questions but do not send your answer to university. It is only for your practice.

UNIT 3 RESEARCH PROBLEM

Structure

3.0 OBJECTIVES

After studying this unit, you will be able to:

- Explain Research Problem and its significance;
- Understand how to identify and formulate research problem;
- Comprehend various steps involved in identification of research problem;
- Explain the formulation of research objectives; and
- Define propositions and convert them into verifiable research hypotheses

3.1 INTRODUCTION

In the previous unit, we discussed the review of literature in detail. Through this process, researchers often identify research gaps—areas that have not been adequately explored or studied. These gaps represent potential research problems that a researcher can focus on. If we discuss in length, it is well versed on the basis of previous units that research is an attempt to answer unexplored questions and helps develop new theories, principles, or generalizations, which open either a new perspective of understanding in the

relevant area of knowledge, or helps in validating the existing knowledge. Research questions require gathering the relevant data, analyzing, and interpreting it in order to reach a conclusion and provide a solution to the research problem.

Figure 3.1: Finding a Research Problem

Research is a gateway to acquire new knowledge that depends on how relevant the problem has been discovered and research questions have been answered. At first, researchers may struggle with 'problem blindness,' where the problem seems too vague or unclear. Therefore, the identification of a problem is the most crucial and difficult step in the research procedure. If this step isn't done carefully, the research process may lose direction. It is a systematic inquiry to get answers to the questions about a phenomenon through the scientific research process.

Problem identification and its formulation is innovatory and individualistic rather than ordinary and mechanical. The identification of a research problem entails a great deal of patience and insightful thinking. In the beginning, a researcher finds identification of a research problem very difficult. This could be due to insufficient knowledge of the research process and lack of preparedness. The Researcher could be unfamiliar with the fields in which scientific enquiry is required and the process he/she is to follow for selecting a suitable area for research.

In this unit, you will learn about the significance of the research problem and various methods of identification of a Research Problem. We will discuss how to formulate research objectives and convert propositions into testable hypotheses. Formulation of Research objectives will also be explained to you in this unit.

3.2 SIGNIFICANCE OF RESEARCH PROBLEM

A well-defined research problem is of utmost importance as it not only provides direction to the study but also determines the research design and regulates the relevance of findings. For instance, a research problem emphasizing the impact of teaching methods on the performance of BBA students provides a clear framework for data collection, determination of sample size, methodology, and potential research impact.

Table 3.1: Poorly Defined Problem vs Well-Defined Problem

Poorly Defined Problem	Well-Defined Problem
Vague and unclear *e.g., "How to improve sales?"*	Specific and focused *e.g., "What marketing strategies increase sales of Product X in Market Y?"*
Lacks measurable objectives	Clear objectives with measurable outcomes
No connection to research design	Guides methodology and data collection

Significance of a research problem are as follows:

1. The research problem is the beginning point of any research process. A well-chosen research problem defines the focus and scope of the research, helping researchers to decide what they want to research and why it is significant.

2. A well-defined research problem elucidates why the research is required, what gaps exist in available literature in the domain, and how the research will advance the knowledge in the domain and provide possible solutions to issues. Hence, it provides the rationale for the undertaken research.

3. The nature of the research problem impacts the selection of research design, tools and techniques of data collection, and selection of data analysis techniques. It assures that the research must be in line with the objectives and the methods, tools and techniques chosen would be appropriate for solving the problem.

4. A well-defined research problem directs the formulation of specific research questions or hypotheses. These research objectives guide the entire research procedure, from literature review to data analysis, assuring that the study remains relevant and focused.

5. The research problem is also significant due to its potential influence. A well-defined problem may lead to findings that are useful, meaningful, and also applicable in the real world situations, practice, influencing policy, or further research.

6. A significant research problem often attracts funding and resources. Research problems addressing a critical need or has the potential to have a significant impact is more likely to get support from the stakeholders.

In summary, the significance of a well-defined research problem lies in its potential to provide a focused, clear, and valid foundation for research, guiding the entire research process and determining the relevance and influence of the research findings.

3.3 FORMULATION OF RESEARCH PROBLEM

Research problem formulation is the initial step of the research process. Research problem is a precise and well-defined problem or question that a researcher aims to investigate through research. It is the beginning point of any research, as it sets the purpose, direction, and scope of the study. Formulation of the research problem involves two main issues i.e., detailed apprehension of the problem and converting the same into meaningful terms from an analytical perspective.

One way of understanding the research problem is to discuss it with experts in the subject or with one's own colleagues. Generally, teachers put forward the broad problem and it depends upon the researcher to narrow it down and express it in operational terms. The research problem is generally determined by the administrative heads in governmental and non-governmental organizations. A researcher discusses the problem with those administrative heads as to how the problem originally arose and what could be the possible solutions for it.

Another way of problem comprehension can be reviewing the available literature in the particular area. The researcher is required to examine all available related literature to get all the available information on the problem. The literature provides information about the variables, concepts, theories, research gaps etc. Through literature review, the researcher gets to know what studies have been already conducted on the subject matter and what were the findings. The outcome enables the researcher to frame his/her own research problem. After framing the problem, expressing the problem in operational terms is done. Defining the research problem is the foundation of the entire research process. The problem statement helps in deciding the data to be collected, features of the data, associations between variables to be examined, and the selection of methods and techniques in these investigations.

In fact, formulation of the research problem follows the following sequential process:

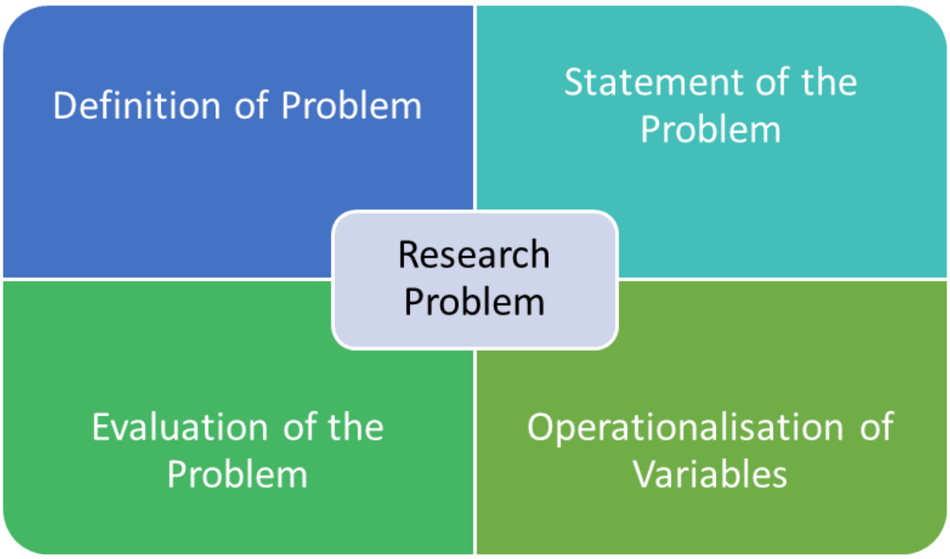


Figure 3.2: Formulation of Research Problem

3.3.1 Definition of the Problem

Once the researcher has identified a research problem, the next step is to define the research problem. The research problem has to be well defined and narrowed to a manageable size. Research questions including sub-questions, scope, and the limitations of the investigation must be determined at this stage. The researcher must formulate his/her overall plan to see if it is feasible. Research problems must be defined in an easy-to-understand manner. A research topic must be based on available factual information, which is used to introduce the topic and become the foundation of research questions. Generally, reviewing the past studies is essential to get an idea about what may be done in future.

3.3.2 Statement of the Problem

A good research problem must state what is to be done or what are the research question(s). It must limit the scope of the study to feasible research questions. In this step, the researcher is required to introduce the topic in detail, its theoretical basis and its principal assumptions, and define the problem in specific and workable questions. All research questions must be related to the problem. Each element should be separated and arranged in a logical sequence under the major divisions.

3.3.3 Operationalisation of Variables

In defining a research problem, the researcher must ensure that it is neither framed so generally as to make it vague, nor stated so narrowly as to make it trivial. In this direction, to decide the variables of the study and their operational definition becomes significant. To illustrate, suppose that you want to study the "Effectiveness of Yoga on the stress of people". This statement is vague and it communicates in a general way what you are planning to conduct. Here, it is necessary to define the problem precisely. For this the initial step is to clearly decide the variables involved in the research problem and define them in operational terms.

The variables involved in the study are "effectiveness" and "stress". These variables must be defined in operational terms. It is important for you to identify what all indicators of effectiveness you will use or how you will measure the presence or absence of the phenomenon expressed by the term "effectiveness". Likewise, you should define the other variable "stress" and identify its indicators that will be used to measure it.

3.3.4 Evaluation of the Problem

When assessing a proposed research problem, an investigator should consider their expertise in the subject, the financial and field constraints, potential challenges with data availability, and time constraints. You should refine a suggested research problem to a more focused one after assessing it. The research topic is formulated by stating certain research questions that you plan to investigate using scientific methodologies.

It would be beneficial to ask yourself a number of questions before you initiate the research. These questions should be thoughtful in the evaluation

of the problem on different criteria. All questions should be answered affirmatively before the study begins. The questions that we should ask are as follows:

i) Is the problem researchable?

There are definite problems that cannot be answered through the research process. A researchable problem is concerned with the association existing between variables which can be defined and measured.

ii) Is the problem new?

The research problems which have already been studied sufficiently are of no use to study further as it leads to duplication. To avoid such duplication, it is essential to explore the available literature in the field carefully. The problem to be investigated should be selected only when you are confident that it is novel which has never been investigated before. However, it should be noted that a researcher may conduct a study again when he/she wants to validate its conclusions or extend the validity of its findings in an entirely different situation.

iii) Is the problem significant?

The research problem must fill the gaps in the existing literature available in the domain. The findings of a study should either become a foundation for a theory, principles or generalizations. Besides, they should lead to original problems for further research and have some useful practical implications.

iv) Is the problem workable for the particular researcher?

a) Research competencies

The problem must be related to the domain in which the researcher is competent and proficient. He/She must have expertise and competencies required to establish and administer the data gathering tools, and analysis of data.

b) Interest and enthusiasm

The researcher should genuinely be interested in and enthusiastic about the problem which he/she wants to study.

c) Financial considerations and feasibility

The research problem should be financially feasible. The investigator has to ensure that they have the funds required to carry out the research. The cost of undertaking a study is a crucial component of feasibility. It is necessary to estimate the cost and assess the availability of funds. This would verify whether the project can actually be executed.

d) Time

Projects are a time bound task. Majority of you, if not all, are already engaged in a number of activities at your workplace, home

and at social organizations. It is necessary to ascertain the time required to accomplish a study, beyond the assessment of total period, it is important to determine the period of the year in consideration of the study.

e) **Administrative considerations**

In addition to financial, personal and time constraints, the investigator should consider the nature of data, specialized personnel, administrative facilities, and equipment needed to accomplish the study successfully. The researcher should confirm whether he/she is able to get the support from various administrative authorities for the collection of various types of data.

3.4 TECHNIQUES INVOLVED IN IDENTIFYING RESEARCH PROBLEM

Identifying a research problem is a complex but an important step in the research process. Following are several techniques of identifying research problem are given below:

1. **Review of Related Literature**: As we have already discussed about the review of literature in the preceding unit. Review of related literature is one of the crucial techniques involved in identifying research problems. Literature comprises all published and unpublished sources of data in the researcher’s area of interest. It may include journals, e-journals, newspapers, Dissertation Abstract International (DAI) database, magazines, reports, doctoral thesis, government publications, etc. Related means the chosen aspect of the field in which one intends to do research. If one wants to do research in the domain of e-marketing, one has to see the existing literature related to marketing approaches. This is the meaning of review of related literature. The advantages of the review of related literature includes: First, it helps in identifying research gap(s) in the existing literature by delineating whether any research has been conducted or not in the specific field of researcher’s interest. If not, it means the gap exists. Second, it provides various perspectives and methodologies to be applied in investigating the problem. Third, it helps in recognizing potential variables that may require more investigation. Fourth, it helps in narrowing the scope of the study into a viable research problem that is significant, relevant and testable.

2. **Unutilised/Underutilised Theory**: Theoretical frameworks are the foundation of research that provides a lens through which phenomena is examined. Theories offer justifications for how and why things happen. It is widely known that there are various theories in different subjects, like Marketing, Physics, Chemistry, Psychology, Mathematics, Education, Economics, etc. are available. The theories are developed so that these can be used by people to make their life comfortable. It is seen when the company's manager publicly praised an employee in a team meeting and that employee received a “Best Performer” award. This recognition boosts morale and motivates the employee to maintain or

even exceed current performance levels. This means recognition motivates individuals. There might be different ways of motivating people to do better. Hence, there is a scope of doing studies related to Theories of Motivation.

3. **Seminars/Conferences**: Seminars and Conferences are organized in all the disciplines. In seminars/conferences the researchers present their original research work related to the theme provided by the seminars/conferences, which is very rich in ideas. One may get an interesting idea related to which the researcher can think of conducting research. Hence, the seminar/conference is an important source of identifying the research problem. After the presentation of research work, the researcher may get feedback related to the presented work. Now he/she can modify the presented paper and send it to some journal. In this way, seminars/conferences may become another source of problems.

4. **Informal Source:** Newspaper, Radio, TV, Podcast, YouTube, Social Media Players etc., are an informal source of a research problem refers to propositions or issues that originate from day-to-day experiences, observations, or discussions as opposed to formal academic or professional literature. These kinds of sources might include conversations with peers, media reports, personal experiences, or even casual readings. These sources may highlight feasible research gaps in understanding or practical problems that require solutions. Let us take an example. Suppose you are a school teacher who regularly talks with other teachers about student engagement in online classes. During these talks, you realize that many school teachers are struggling to keep their students interested in online classes. This may lead you to identify a research problem: "What techniques can boost student engagement in virtual classrooms?" One should understand that these sources are not a very good source of research problems. One may use it to conduct research for writing papers but not for research leading to any degree.

3.5 IDENTIFICATION OF RESEARCH PROBLEM

The research problem identification starts with the decision maker and some dilemma that might be faced by him/her. This action-oriented problem addresses the issue of what the decision maker must do. Sometimes, this might be associated with actual and immediate difficulties encountered by the manager or gaps confronted in the existing body of knowledge. The wide decision problem has to be narrowed down to an information oriented problem which emphasizes on the data required to arrive at any significant conclusion. In table 3.2, a set of decision problems and succeeding research problems addressing them are given.

Table 3.2: Decision Problem vs. Research Problem

Decision Problem	Research Problem
What should be done to increase the sales of natural products in the foreign market?	What is the perception and purchase intention of health-conscious consumers for natural products in the international market?
How to improve the retention rates in the BPO sector?	What is the impact of increment and work environment on retention intentions of the BPO employees?
How to improve customer satisfaction with our online services?	What is the impact of user experience design and functionality on customer satisfaction?
What should be the effective strategy for entering a new international market?	How do consumer preferences and buying behaviors vary in the target market compared to existing markets?
Why are our projects consistently experiencing cost overruns?	How do project management practices affect cost control?
Why is the company's market share declining?	What are the factors contributing to the decline in market share?

A researcher should follow the following major tasks in evaluating a problematic situation:

i) Accumulate the facts and information that might be related to the problem,

ii) Determining the relevance of the facts through observation,

iii) Finding any association between facts that might reveal the key difficulty,

iv) Present various explanations for the cause of difficulty,

v) Determine whether these explanations are relevant to the problem through observation and analysis,

vi) Trace the relationship between explanations that may provide an insight into the problem solution,

vii) Find the relationship between facts and explanations, and

viii) Questioning the assumptions underlying the analysis of the problem.

A researcher may face problems in administration of a college, in classrooms and in all other fields of education such as the teaching-learning process, textbooks, curriculum, physically and mentally challenged children, guidance and counseling, etc. A research problem must be firmly rooted in knowledge. The investigator should first identify a general area of research in a particular domain and then explore the available literature in that field.

The identification of a problem follows the following procedure:

- Understanding the facts, theories and available information in the domain. The research focus is improved by what he already knows, what all studies in that field have been conducted and what is required to be discovered.
- The research problem may arise from the natural interest of a researcher, or from the researcher's curiosity about something observed, sensed, or thought about.
- Life situations, implications advanced by technological advancements and relationships established by related researchers constitute the problem situation.
- Future scope section in the research reports provides the gaps that exist in the domain, helpful in keeping the researcher informed about what studies have been already done and what could be done in future research studies.
- New knowledge in the area of researcher's interest emerging through the new arrivals of books, researchers and journals extend a situation for further research.

Check Your Progress A

1) Distinguish between decision problem and research problem.

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

2) What are the various techniques of identifying research problems?

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

3) What do you understand by evaluation of the problem?

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

4) What is the role of internal sources in identifying a research problem?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

5) Fill in the blanks:

a) Research problem is a _______ and well-defined problem or _______ that a researcher aims to investigate through research.

b) The identification of a problem is the most _______ and _______ step in the research process.

c) A well defined research problem directs the formulation of specific research _______ or _______ .

d) Identifying _______ contradictions or inconsistencies in the available literature can be the foundation for formulating a research problem.

e) The research problem must fill the gaps in the _______ literature available in the field.

f) A research problem must be _______, meaning that it can be examined within the time available, resources, and expertise of the researcher.

g) Conducting a thorough _______ review is a common technique for identifying a research problem, which helps to identify gaps in the existing knowledge.

3.6 SELECTING A RESEARCH PROBLEM

Formulation of research problems is of great importance as it determines the quality and validity of the components of a research report. It is therefore important to allocate ample time to the formulation of research problems as better and rationally formulated research problems lead most inevitably to clearer research and more consistent output and progress. The key sets of factors that have been identified as being involved in the formulation of research problems are the internal and external factors or criteria. The internal factors can be defined as those factors that are determined by the researcher, such as the interest of the researcher, the researcher's own resources, and the researcher's competence and expertise. External factors are those which cannot be determined by the researcher and include problem feasibility, originality of research problem, facilities, usefulness, research personnel and social relevance.

Internal Factors:

1. **Interest:** Interest of the researcher is considered as an important factor which directs the formulation of research problems. As research is a time consuming process and demands a lot of hard work, it facilitates if the

researcher selects a topic which interests him/her. Otherwise, the researcher may find it difficult to establish and maintain the required levels of enthusiasm and commitment. Interest in a problem is usually driven by the researcher's educational background, outlook, experience and sensibility.

2. **Expertise:** It refers to the ability of the researcher to design and conduct research (including data collection, data analysis and interpretation, etc.). Having interest in a problem is insufficient. A researcher must be competent enough to plan and conduct a study. He/she must possess adequate knowledge and expertise of the subject-matter, methodology and statistical procedures.

3. **Researcher's own resource:** Studies which are funded by researchers require to take into consideration their financial capacity is quite important. The researcher will not be able to accomplish the research work if it is beyond the researcher's financial capability except if he/she gets supported financially. The ability to allocate adequate time and resources must be carefully considered while formulating a research problem as it is a time consuming procedure.

External Factor:

1. **Data availability:** If a study requires collection of relevant information or data (journal, proceedings, reports), then it becomes important to ensure that these materials must be available and in the proper format.

2. **Ethics:** One should consider the ethical issues while formulating a research problem. Often the targeted population might be adversely affected over the research period. In ICT, some situations may arise, especially related to collected information security, which may concern certain authorities. Thus, it is always crucial to identify ethics-related issues during the formulation of research problems.

3. **Relevance:** It is significant to always select a topic that is relevant to one's interest and profession. It is important to ensure that one's study contributes to the existing body of knowledge.

4. **Novelty of the problem:** The research problem must be original. Studying a problem which has already been done by others is of no use and it is a waste of one's time and energy.

5. **Research ability of the problem:** The problem must be researchable, i.e., amenable to discovering answers to the research questions involved through the scientific method.

6. **Importance and urgency:** Problems that require investigation are infinite, but available research efforts are very much limited.

7. **Feasibility:** A problem may be original and also important, but it cannot be selected if research on it is not feasible.

8. **Facilities:** Research requires certain facilities, such as data analysis facility, suitable and competent guidance, well equipped library facility,

etc. Thus, the availability of the facilities pertinent to the problem should be considered.

9. **Research personnel:** Research undertaken by research organizations and academicians require research related services from research officers and investigators. In developing countries, research is still not a prospective career. Thus, talented people are not attracted to research projects.

10. **Usefulness and social relevance:** Above all, the research must contribute to the existing body of knowledge or provide a solution to some significant practical problem. Hence, It should be socially relevant.

3.7 FORMULATION OF RESEARCH OBJECTIVES

Research objectives refer to the specified aims or goals of a study. They provide a clear and brief description of what the researcher aims to achieve by conducting the study. Generally, the objectives are based on the research questions and hypotheses developed at the beginning of the research study and are used to direct the research process. Writing clear and precise research objectives is a crucial part of any research project, as it helps to direct the study and assures that it is focused and relevant. Following are the steps one needs to follow while formulating the research objectives:

- **Identify the research problem:** A Researcher needs to identify the research problem that he/she wants to study before writing research objectives. The identified problem must be clear and specific that can be undertaken through research.

- **Define the research questions:** Formulation of research questions based on identified research problems. These research questions must be specific and provide direction to the research process.

- **Identify the variables:** After defining research questions, a researcher needs to identify the variables that he/she will be studying. These are the key factors that will be manipulating, measuring, or analyzing to provide answers to specific research questions.

- **Write specific objectives:** Writing specific and measurable research objectives help to answer research questions. These objectives must be clear and brief and should indicate what a researcher aims to achieve through this research process.

- **Use the SMART criteria:**

Figure 3.3: SMART criteria for formulating Research Objectives

To assure that research objectives are well-defined and attainable, use the SMART criteria. This means that formulated research objectives must be Specific, Measurable, Achievable, Relevant, and Time-bound. Using the SMART criteria is a key step to formulate strong, clear and easier to understand research objectives. Hence, make sure that research objectives meet the following criteria:

- **Specific:** Research objectives must be precisely written and leave no room for confusion. The researcher must be specific about desired outcomes of the study undertaken. This can facilitate him/her keeping them narrow and focused.
- **Measurable:** The research objectives must be measurable in order to achieve them. A researcher can develop metrics to measure the progress toward achieving research objectives.
- **Achievable:** Research objectives must be realistic to achieve and a researcher must make sure that he/she has adequate resources and budget to attain the objectives.
- **Relevant:** A researcher must ensure that research objectives are relevant to research and overall objectives. This helps him/her to stay motivated and on track throughout the study.
- **Time-based:** To achieve the research objectives timely, a researcher may decide a deadline to facilitate keeping him/her on track throughout the research process. The investigator can set a major deadline for the entire research project and smaller deadlines for each objective.
- **Revise and refine:** After formulation of research objectives, revising and refining them becomes important to ensure that they are achievable, clear, and concise. Make sure that objectives are in accordance with research questions and variables, and they will help to provide answers to the identified research problem.

In summary, formulating clear and concise research objectives is important as they establish the scope and depth of the study, helps in developing research design, selecting statistical techniques and indicates how the research will contribute to the existing body of knowledge.

3.8 PROPOSITION AND HYPOTHESES

A proposition is a tentative and hypothetical relationship between constructs that is formulated in a declarative form. For example, "An increase in student hard work causes an increase in their academic performance." This declarative statement can be true or false, but must be empirically verifiable using collected data, so that we can determine whether it is true or false. In general, Propositions are derived based on either logic (deduction) or empirical observations (induction). As the propositions are relationships between latent constructs, they cannot be measured directly. Instead, they are tested by examining the relationship between measures (variables) corresponding to those constructs. When a proposition can be examined for its validity through empirical investigation, then it is called a hypothesis. It is

a condensed generalization. The variables are generally used in hypotheses to delineate the attributes or characteristics that need to be measured. Since study hours and grade point average are the operational measures of hard work and academic performance respectively, the above proposition can be stated in the form of the hypothesis: "An increase in students' study hours causes an increase in their grade point average in exams." Propositions are stated in the theoretical plane, whereas hypotheses are stipulated in the empirical plane. Thus, hypotheses are empirically verifiable using collected data, and may not be accepted if not supported by empirical observations. Certainly, the aim of hypothesis testing is to deduce whether the corresponding proposition is valid or not.

Hypotheses can be strong or weak. "Students' study hours are associated with their academic performance" is an example of a weak hypothesis, as it signifies neither the direction of hypothesis (i.e., whether the association is positive or negative), nor its causality (i.e., whether hard work causes academic performance or academic performance causes hard work). A stronger hypothesis is "students' study hours are *positively* related to their academic performance", which specifies the directionality but not the causality. A better hypothesis is "students' study hours have a significant positive influence on their academic performance", which states both the directionality and the causality (i.e., hard work causes academic performance, and not the reverse).

Also note that scientific hypotheses must clearly specify independent and dependent variables. In the given hypothesis, "students' study hours have a significant positive influence on their academic performance", it is clear that hard work is the independent variable i.e., the cause and academic performance is the dependent variable i.e., the effect. Furthermore, it is also understood that this hypothesis can be determined as either true (if more number of study hours leads to higher academic performance) or false (if more number of study hours has no significant positive influence on academic performance). Later on in this book, we will learn how to empirically test the hypotheses.

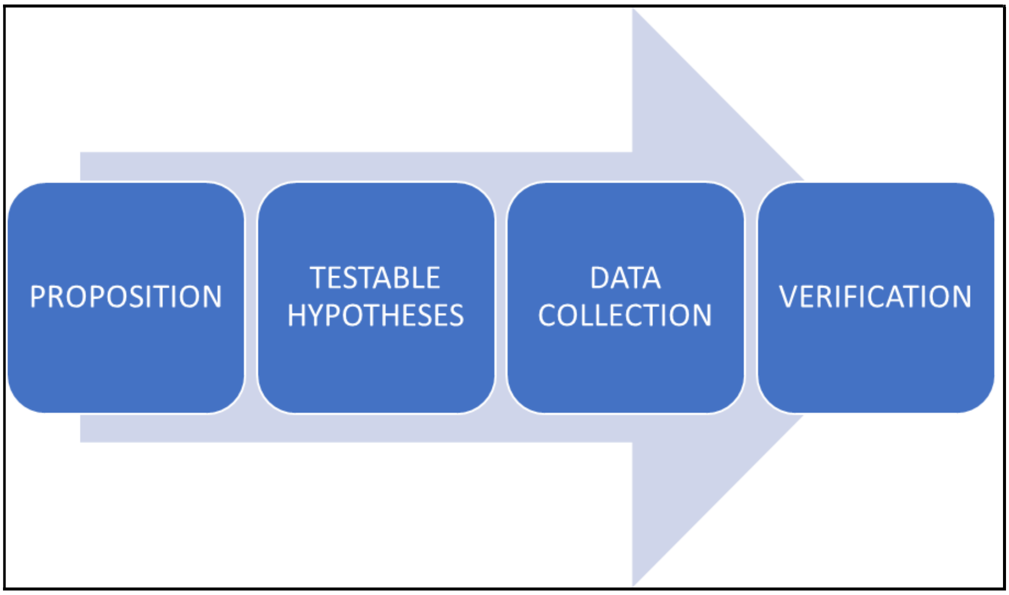


Figure 3.4: Steps of Hypothesis Formulation and Testing

Check Your Progress B:

1) What are the different criteria for selecting a research problem?

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

2) How do researcher's interest and expertise help in selecting a research problem?

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

3) What is the social relevance of a research problem?

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

4) What do you understand by a feasible research problem?

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

3.9 CASELETS

Caselets are concise scenarios that present specific problems along with relevant data. They are commonly used in research to simulate real-world challenges, enabling critical thinking, data analysis, and problem-solving.

Caselets serve as a practical tool to explore research problems by combining contextual information with data interpretation. Each caselet is designed to highlight a unique problem, prompting logical reasoning and analytical skills to find solutions.

Key features of Caselets

Following are the main features of caselets:

Contextual Information

Caselets begin with a brief description of a situation or challenge. These scenarios are often inspired by real-life contexts from industries like healthcare, education, business, or technology.

Data Presentation

Data is provided in formats like tables, charts, or figures. This data serves as the foundation for identifying patterns, trends, and key insights.

Research Problem Identification

Each caselet centres around a well-defined research problem. This problem encourages analysis and a deeper understanding of the factors influencing the situation.

Critical Thinking and Analysis

Caselets require logical reasoning to combine insights from the given scenario and data. This approach fosters critical thinking and structured problem-solving.

Practical Relevance

Caselets reflect real-world issues, offering valuable experience in addressing complex research problems effectively.

Caselet 1: Healthcare and Technology

Figure 3.5: Healthcare and Technology

Scenario: A startup in the healthcare industry has recently developed a new wearable gadget designed to monitor key signs and alert users to potential health problems before they become severe. Initial Trials have shown positive results, but the company is struggling to recognise why the device performs inconsistently across different demographic cohorts.

Research Problem: How can the healthcare startup identify and address the factors which lead to the inconsistent performance of their wearable gadget among different demographic cohorts?

Caselet 2: Education and Learning Outcomes

Figure 3.6: Education and Learning Outcomes

Scenario: A school division has executed a new digital learning platform aimed to enhance students' engagement and improve academic performance. Initial report suggests mixed results, with some students progressing while others show decreased performance or no improvement in the performance.

Research Problem: What factors are causing the varied effectiveness of the digital learning platform, and in what ways can the district best optimize its implementation to benefit all the students?

Caselet 3: Business and Consumer Behavior

Figure 3.7: Business and Consumer Behavior

Scenario: A manufacturing company has launched a new line of environment-friendly products, however initial sales figures are lower than anticipated. Market research stipulates that consumers are interested in sustainability, though they are not purchasing the new eco-friendly products.

Research Problem: Why are consumers not buying the new environment-friendly products despite being interested in sustainability, and what strategies can the company implement to boost sales?

3.10 LET US SUM UP

In this Unit, we have discussed the significance of research problem, its formulation and techniques of identifying research problem, which includes review of related literature, unutilised or underutilized theory, seminars/conferences and various informal sources.

Further we described the procedure to be followed for identification of research problem flowed by selection of a feasible research problem. There are different internal and external factors concerned for identifying a research problem. Internal factors can be defined as those determined by the researcher and external factors are those that are not determined by the researcher.

In addition, we discussed the steps one needs to follow while formulating the research objectives. Using the SMART criteria is a key step to formulate strong, clear and easier to understand research objectives. To assure that research objectives are well-defined and attainable, one should use the SMART criteria. This means that formulated research objectives must be Specific, Measurable, Achievable, Relevant, and Time-bound.

A proposition can be defined as a tentative and hypothetical relationship between constructs that is formulated in a declarative form. When a proposition can be examined for its validity through empirical investigation, then it is called a hypothesis. It is a condensed generalization.

3.11 KEYWORDS

Hypothesis: When a proposition can be examined for its validity through empirical investigation, then it is called a hypothesis.

Proposition: A proposition is a tentative and hypothetical relationship between constructs that is formulated in a declarative form.

Research Problem: Research problem is a precise and well-defined problem or question that a researcher aims to investigate through research. It is the beginning point of any research, as it sets the purpose, direction, and scope of the study.

Research Objectives: Research objectives refer to the specified aims or goals of a study. They provide a clear and brief description of what the researcher aims to achieve by conducting the study. Generally, the objectives are based on the research questions and hypotheses developed at the beginning of the research study and are used to direct the research process.

Review of Related Literature: Review of related literature is one of the crucial techniques involved in identifying research problems. Literature comprises all published and unpublished sources of data in the researcher's area of interest. It may include journals, e-journals, newspapers, dissertation

abstract international (DAI), magazines, reports, doctoral thesis, government publications, etc.

SMART criteria: SMART criteria is a key step to formulate strong, clear and easier to understand research objectives. This means that formulated research objectives must be Specific, Measurable, Achievable, Relevant, and Time-bound.

3.12 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress A:

5. a) precise, question b) crucial, difficult c) questions, hypotheses d) theoretical e) existing f) feasible g) literature

3.13 TERMINAL QUESTIONS

1) Define research problem. What is the significance of a research problem?
2) Briefly explain the process of formulation of research problem.
3) Describe the various techniques involved in identifying a research problem.
4) Explain the procedure of identification of research problem.
5) Briefly explain the internal and external factors of selecting a research problem.
6) Define research objectives. What are the steps of formulation of research objectives?
7) Briefly explain SMART criteria.
8) Explain proposition and hypothesis in research with suitable examples.
9) What are the major tasks in evaluating a problematic situation?

3.14 FURTHER READINGS

1. "Research Design: Qualitative, Quantitative, and Mixed Methods Approaches" by John W. Creswell
2. "Research Methodology: Methods and Techniques" by C.R. Kothari
3. "Research Methodology: A Step-by-Step Guide for Beginners" by Ranjit Kumar
4. "Understanding Research Methods: An Overview of the Essentials" by Gerald R. Lynch
5. "Research Methods for Business Students" by Mark Saunders, Philip Lewis, and Adrian Thornhill

Note	These questions are helpful to understand this unit. Make an effort to write the answer to these questions but do not send your answer to the university. It is only for your practice.

UNIT 4 RESEARCH DESIGN

Structure

4.0 OBJECTIVES

After studying this unit, you will be able to:

- Define research design and understand the significance of research design;
- Categorise and explain various research design;
- Compare and contrast the basic research design: exploratory, descriptive, and experimental;
- Understand how respondents affect research design choices; and
- Apply appropriate research designs to specific business scenarios.

4.1 INTRODUCTION

In the preceding unit, we discussed the research problem. Once the research problem is determined, the next stage is to choose the design for reaching the research goals in the most effective way. This unit explains the definitions and categories of research designs. Research designs are mainly divided into two categories: exploratory research design and conclusive research design. Conclusive research design is further categorised into two categories: descriptive research design and experimental research design. This unit will provide a detailed discussion of both types. Next, we will also look at the two

types of descriptive research designs i.e, cross-sectional research design and longitudinal research design.

Research design involves the strategic planning of how to do research. The strategy includes determining what will be observed, how it will be observed, when and where it will be observed, why it will be observed, how to record the observations, how to analyse and interpret the findings, and how to draw broad conclusions. Research design is a comprehensive blueprint outlining the specific strategies and methods that will be used to accomplish the objectives of a research project. A research design is as important to a study as a blueprint is to building construction.

4.2 DEFINITION OF RESEARCH DESIGN

A research design is a structured plan that guides a study. It outlines the steps needed to collect the data required to solve a research problem. Even if the overall direction of the study is already clear, the research design focuses on the specific steps needed to achieve the study's goals. It forms the foundation for conducting research, and the success of any business research project depends on having a well-thought-out research design.

Jenkins-Smith et al. (2017) define research design as the steps taken to collect and analyze research data. It combines both research methodology and research methods. Similarly, Creswell and Creswell (2018) describe research design as a framework that connects research questions, data collection, and analysis, ensuring meaningful results. Saunders, Lewis, and Thornhill (2019) call it a strategic plan that ensures the research process is practical and aligned with its objectives. Bell, Bryman, and Harley (2022) define it as the structure for organizing research activities to address specific questions, ensuring the methods used are appropriate. Bhattacherjee (2023) highlights research design as a blueprint that links theoretical concepts to observations.

For instance, in a study on user satisfaction with an online service platform, the research design might include identifying factors like ease of use, response time, and overall satisfaction. The researcher would then select a diverse group of users, gather data using surveys, interviews, or app usage tracking, and analyze the results to identify patterns, such as features that users find most beneficial or areas needing improvement. This example shows how research design aligns all steps in a study to achieve meaningful results.

Key Elements of Research Design

Research design often includes the following activities:

- Identifying the required information.
- Deciding if the design will be experimental, descriptive, or exploratory.
- Determining the sequence of procedures for data collection or measurement.
- Designing and testing tools like questionnaires for data collection.

- Selecting the sample size and sampling method (qualitative and/or quantitative).
- Planning strategies for analysing the collected data.

Distinction Between Research Design and Research Method

It is essential to differentiate between research design and research method. Research methods refer to the processes used to collect data, while research design provides the overall framework that outlines how the research question will be addressed.

Thus, research design is a critical part of the research process, guiding every step. It specifies how the study will be conducted to answer the research question effectively.

4.3 SIGNIFICANCE OF RESEARCH DESIGN

The research design is essential for ensuring the dependability and accuracy of research results, as it serves as the framework for the whole research procedure. Just like building a home requires a well-thought-out blueprint by an architect, a research project needs a well-planned design. This design must be finalised before gathering and analysing data. An organised and well-planned research design ensures that different research activities run smoothly, allowing for a more efficient research process that provides in-depth insights while minimising the use of time, efforts, and resources.

Reliability refers to the consistency and stability of research findings. In other words, if the research were to be conducted again under similar circumstances, reliable results would provide the same conclusions. Validity, however, assesses whether the research really investigates the subject matter it aims to analyze. An effective research design guarantees the accuracy of the results and their alignment with the real-world scenario targeted by the study. Reliability and validity are also essential for making trustworthy results, and a robust research design addresses both reliability and validity by reducing mistakes, biases, and inconsistencies.

Take, for example, a business planning to study customer preferences before launching a new product. If the research design is poorly planned, it might lead to biased data, irrelevant results, or incorrect conclusions, causing the business to make poor decisions. On the other hand, a well-structured research design ensures that the right sampling methods are used, relevant data is collected, and proper analysis techniques are applied. This helps the business gather useful insights, align the product features with customer preferences, and reduce the risk of a failed product launch. **Reliability** in this context means that if the company conducts the survey multiple times, they should consistently get similar results, confirming that their findings are stable and dependable. On the other hand, **validity** ensures that the survey actually measures what it intends to—customer preferences—rather than unrelated topics.

An effective research design seamlessly incorporates several components. The theoretical and conceptual frameworks must be in perfect harmony with the research aims. Similarly, the methods used to gather data should align with the research's objectives, theoretical framework, and analytical methodology. Thorough planning of the research design is essential, since even little mistakes may put the whole project's success at risk. Researchers benefit from the process of refining their ideas and identifying any problems before collecting data, which ultimately improves the quality and dependability of the study.

The need for research design includes:

- **Reducing inaccuracies**: It ensures systematic planning and execution, reducing the chances of errors.
- **Enhancing efficiency and reliability**: Effective research design streamlines research processes, yielding dependable results.
- **Eliminating biases and marginal errors**: It incorporates measures to control and mitigate potential biases.
- **Minimising time wastage**: Good research design organises research activities to optimise the use of available time.
- **Facilitating material collection**: It guides in gathering relevant literature and resources.
- **Testing hypotheses effectively**: It provides a solid foundation for hypothesis testing.
- **Resource planning**: It offers insights into required resources like money, manpower, time, and effort.
- **Expert review**: It provides a clear overview for peer feedback and validation.
- **Guidance in the right direction**: It keeps the research focused and aligned with objectives.

In business research, where decisions rely heavily on accuracy of findings, a robust research design is for achieving meaningful, actionable, and reliable results that stand the test of both reliability and validity.

4.4 TYPES OF RESEARCH DESIGN

The researcher has several designs to choose from to explore the research objectives. These designs can be classified in different ways. The most common and easy-to-understand classification is based on the purpose of the study. Figure 4.1 shows a simple classification based on research needs, from general and flexible to specific and more structured approaches. This illustration shows the two types of research, ‘exploratory’ and ‘conclusive’ as separate design options, with subcategories in each.

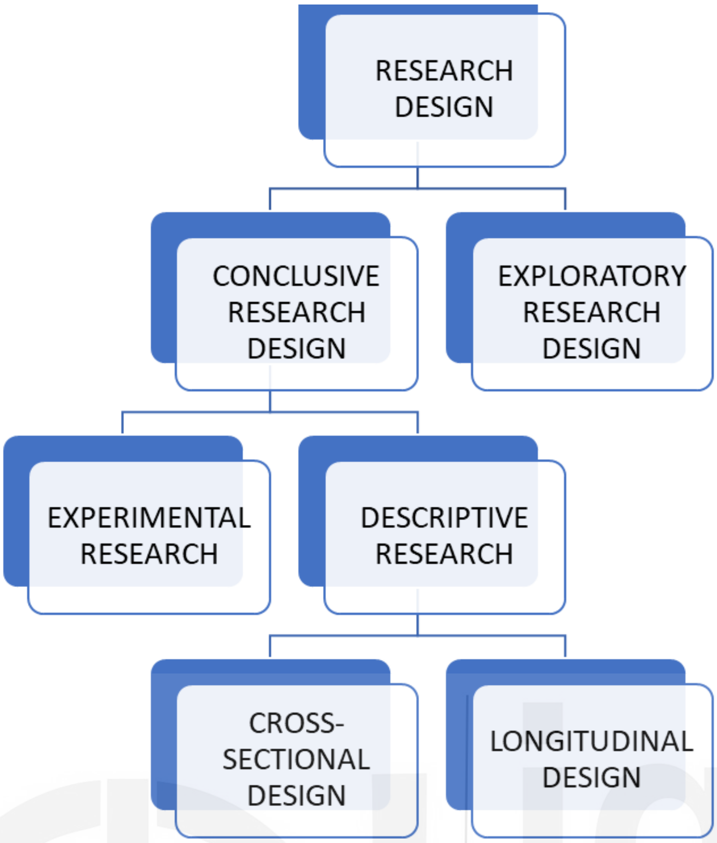


Figure 4.1 Types of Research Designs

4.4.1 Exploratory Research Design

Exploratory designs are the simplest and most flexible research methods. The main purpose of exploratory research is to understand and gain insights into business phenomena. It is often used when the research topic is difficult to measure or when standard measurement methods don't fully capture certain aspects. For example, if a researcher wants to define the 'ambience' of a restaurant, exploratory research can help identify factors like the role of music type, volume, furniture style, colours, lighting, and architectural features. These factors vary and might not be easy to measure, as different customers may associate 'ambience' with different elements. Quantifying these factors might not fully capture the overall atmosphere from a creative perspective. Exploratory research is also useful for better defining a problem, suggesting possible solutions, or gaining deeper insights before using a more conclusive design to confirm the findings.

Figure 4.2 Exploratory Research

Under such conditions, research questions will be used to specify the information required at an exploratory stage rather than forming precise hypotheses or measurements. The chosen research methodology is defined by its adaptability, loose structure, and, in some cases, evolutionary nature. One way to implement a flexible, loosely organised, and evolving method is to interview industry experts in person. The sample is usually small and non-representative since it is selected for maximum insight possible. But the selection process places more emphasis on "quality" people—those who are prepared to be imaginative, innovative, and possibly even to disclose sensitive ideas and actions. Their degree of experience may also determine "quality"; for instance, the number of top executives in Indian airline businesses may be relatively small. It could be possible to get insights even if a small sample of six chief executives from the biggest and fastest-growing airlines gave access to a marketing researcher and disclosed their opinions and behaviors. An exploratory design can be very helpful because it is flexible in terms of the issues to discuss, loosely structured in terms of how probes and additional issues emerge, and evolutionary in terms of who to talk to and the best context in which to gain their confidence and get them to express what they really feel.

In short, exploratory research is valuable when the researcher doesn't have enough information to move forward with the project. Unlike formal research, it is characterised by flexibility and variety in methods. It rarely involves large samples, probability sampling, or structured surveys. Instead, researchers stay open to new ideas as they emerge and may shift focus if something interesting is discovered. They follow this new direction until all possibilities are explored or another path appears. Because of this, the focus of the study can change frequently as new information comes up. Creativity and imagination are essential in exploratory research. The use of exploratory research design is summarised in Table 4.1.

Table 4.1: Summary of Application of Exploratory Research Designs

Understanding the Unknown: Exploratory research design is used to gather background information when little to nothing is known about a problem area.
Problem Definition and Hypothesis Formation: It helps in fully defining problem areas and formulating hypotheses for further investigation or quantification.
Concept Identification and Exploration: Exploratory research design assists in identifying concepts and exploring ideas during the development of new products or business strategies.
Preliminary Screening in Product Development: This research design utilised in early screening processes, such as in new product development, to narrow down a broad range of potential projects to a smaller, more feasible selection.
Understanding Belief and Attitude Structures: It is used to comprehend the structure of beliefs and attitudes, which aids in interpreting data structures in multivariate analyses.

Exploring Underlying Reasons for Statistical Differences: It investigates the reasons behind statistical differences between groups identified through secondary data or surveys.

Addressing Sensitive Issues: It helps to explore sensitive issues from the perspectives of both respondents and interviewers.

Exploring Deeply Held Beliefs and Emotions: It is useful for probing deeply held beliefs or emotions that respondents may find difficult to articulate or rationalise.

Check Your Progress A

1) What is the basic nature of research designs?

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

2) Define exploratory research design.

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

3) What is the significance of research design?

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

4) What do you mean by reliability and validity in research?

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

4.4.2 Conclusive Research Design

As the name implies, conclusive research is designed to deliver definitive insights that aid in decision-making or drawing conclusions. This type of research is typically quantitative, involving numerical data that can be measured, analysed, and summarised effectively. The primary aim of conclusive research design is to provide a detailed description of certain events, to test specific hypotheses, and to thoroughly analyse specific relationships.

Conclusive research design is often characterised by a higher level of precision and structure compared to exploratory research. The study relies on extensive, representative samples, and the collected data undergoes quantitative analysis.

For example A company planning to launch a new smartphone wants to determine which features customers value the most. After conducting exploratory research, such as focus groups and interviews, the company designs a detailed survey for a larger sample of its target audience. The survey includes specific questions asking respondents to rate features like battery life, camera quality, design, and pricing on a scale of importance.

The collected data is analyzed quantitatively to confirm which features are prioritized by customers. Based on these findings, the company makes final decisions about the smartphone's design and features. This example highlights how conclusive research design is used to gather reliable and specific information that aids in making data-driven decisions.

Conclusive research designs can further be classified as either descriptive or experimental. Similarly, descriptive research designs can be classified as either cross-sectional or longitudinal. Each of these classes is explained further in this unit. The application of conclusive research design is summarised in Table 4.2.

Table 4.2: Application of Conclusive Research Design

Profile Analysis: It provides an overview of the distinctive attributes of relevant groups in a business, such as customers, salespeople, organisations, or market segments.
Behaviour Estimation: It calculates the proportion of a certain population that exhibits specific behaviours.
Event Frequency Tracking: A conclusive research design is used to measure the frequency of occurrences of an event. It helps in identifying trends in customer behaviour.
Market representation: It helps in measuring marketing phenomena in order to draw conclusions that may be applied to broader populations or specific target markets.
Data Integration: It enables the seamless amalgamation of information from many sources, particularly in the realms of marketing intelligence and decision support systems.
Perception Assessment: It measures customer perceptions on the attributes of a product or service.
Longitudinal Comparisons: Conclusive research design analyses data over a period of time to observe and measure changes in important occurrences.
Standardised Measurement: It offers uniform and universally applicable metrics for assessing a phenomenon.
Correlation analysis: Conclusive research design is a method used to ascertain the link and strength of correlation between different factors of a phenomenon.
Predictive modelling: It allows for the creation of precise forecasts based on data.

The differences between exploratory and conclusive research designs are summarised in Table 4.3.

Table 4.3: Differences between exploratory and conclusive research

Basis	Exploratory	Conclusive
Objectives	● To gain insights and understanding of a phenomenon. ● To explore and clarify concepts	● To test specific hypotheses and examine relationships ● To measure and quantify findings
Characteristics	● The required information is often loosely specified. ● Flexible and evolving research process. ● Small, non-representative samples. ● Data analysis may be categorised as either qualitative or quantitative.	● Clearly defined information needs ● Formal and structured research process ● Large, representative samples ● Data analysis is predominantly quantitative.
Findings	● May feed into conclusive research ● Can provide context or highlight areas for conclusive studies	● May refine or validate exploratory research ● Can set a context for exploratory findings
Methods	● Expert Interview ● Focus group interview ● Unstructured observations ● Secondary data	● Survey ● Structured observations ● Experiments ● Secondary data

4.4.2.1 Descriptive Research Design

A descriptive research design is a type of conclusive research design that aims to describe a phenomenon under study. A significant difference between exploratory and descriptive research is that descriptive research involves the initial development of particular research questions and hypotheses. Therefore, the required information is explicitly specified. Descriptive research is characterised by being pre-planned and systematic. It is usually derived from extensive and representative samples. A descriptive research design outlines the procedures for choosing the sources of information and collecting data from those sources. The objectives of the descriptive research design are as follows:

- Provide an elaborate depiction or description of the population of individuals being examined. To comprehend the target audience, it may be necessary to conduct a systematic analysis of the information. To

effectively create an advertising and sales promotion campaign for high-end mobile phones, a marketer would need a comprehensive profile of the demographic that purchases luxury items. Therefore, a descriptive study that collects data on the individuals, items, timing, locations, reasons, and methods of purchasing luxury brands would be the required design to achieve the research goals.

- Studies are conducted to assess the coexistence of certain phenomena or factors. For instance, a researcher interested in determining the connection between market volatility and investing patterns may conduct a descriptive study to determine the association between these two variables.

Some examples of descriptive studies in business research:

- **Market Analysis**: A descriptive research design helps to describe market size, consumer buying power, availability of distributors, and consumer profiles.
- **Market Share Analysis**: It helps to determine the percentage of total sales captured by a company and its competitors.
- **Sales Analysis**: It helps to examine sales by geographic region, product line, account type, and account size.
- **Product Usage Analysis:** A descriptive study describes consumer consumption patterns.
- **Advertising Research:** Descriptive studies helps to examine media consumption habits and audience profiles for specific TV programs and magazines.

Descriptive research can be further classified into **cross-sectional** and **longitudinal research,** which will be discussed later in this unit.

4.4.2.2 Experimental Research Design

In general, an experiment is employed to infer a causal relationship. In an experiment, a researcher actively manipulates one or more causal variables and measures their impact on the dependent variable of interest.

Experimental research design is a conclusive research design that is primarily designed to gather evidence regarding the causal relationship between a phenomenon and its consequences. In business research, the management of a firm must consistently make decisions based on presumed causal relationships. These assumptions may not be reasonable, and the validity of the causal links should be evaluated by proper experimental research. For instance, the widely held belief that a reduction in price would result in more sales and a larger market share is not applicable in certain competitive contexts.

Experimental research is suitable for the following objectives:

- To establish the causal relationship between a phenomenon and factors by identifying the independent variables (causes) and dependent variables (effects).
- When the purpose is to ascertain the nature of the link between the causal factors and the result that has to be anticipated.
- When the hypotheses need to be tested (We will discuss this later in this course).

Before we proceed further, learners must know some important concepts used in experimental research design:

- **Independent variables**: Independent variables are also known as explanatory variables or treatments. These variables are manipulated by the researchers, and their effects are measured and compared. For example, the price of a product is manipulated by the researcher to measure its effect on sales. So, here, price is an independent variable.
- **Test Units:** These are individuals, organisations, or other entities being studied to observe their responses to independent variables or treatments. For instance, if a customer's reaction is monitored when the price of a product is adjusted, the customer serves as the test unit in that scenario.
- **Dependent variable:** Dependent variables measure the effect of treatment (independent variable) on the test units. For example, the price of a product is manipulated by the researcher to measure its effect on sales. So, here, sales are a dependent variable.
- **Extraneous variables**: These are any factors other than the independent variables that can influence the test units' responses. These variables can interfere with the experiment by affecting the dependent variable in ways that could weaken or distort the results. For example, the price of a product is manipulated by the researcher to measure its effect on sales, but the sales are also affected by competitor advertising policy. So here, a competitor's advertising policy is an extraneous variable, which is not an independent variable but it affects the dependent variable that is sales.
- **Experiment**: An experiment is established when the researcher manipulates one or more independent variables and evaluates their impact on one or more dependent variables, while controlling the influence of extraneous variables. In our example, price is an independent variable, sales is a dependent variable, and competitor's advertising policy is an extraneous variable.

Experimental research necessitates a structured and organised approach, similar to descriptive research. Descriptive research is not suitable for the examination of causal relationships, despite its ability to ascertain the extent of association between variables. An experimental research design is necessary for this type of investigation, which involves manipulating the causal or independent variables in a relatively controlled environment. An environment of this nature is one in which the other variables that may impact the dependent variable are as closely monitored or controlled as

feasible. The causality of this manipulation is then inferred by measuring the impact on one or more dependent variables.

Before moving forward, it's important for the learner to understand the relationship between exploratory, descriptive, and experimental research. Although exploratory, descriptive, and causal research are considered major types of research designs, the boundaries between them aren't always clear. A single business research project can involve multiple types of research designs, serving different purposes. The choice of which research designs to use depends on the specific problem being studied. Refer to Table 4.4 regarding guidelines for choosing different research designs.

Table 4.4: Guidelines for choosing exploratory, descriptive, and experimental research designs.

Exploratory research is best used when little is known about the situation. It's suitable in the following cases: a) When the topic is difficult to measure in a structured, quantifiable way. b) When the problem needs a clearer definition. c) When alternative actions need to be identified. d) When research questions or hypotheses need to be developed. e) When key variables need to be identified and classified as dependent or independent.
Exploratory research is often the starting point in a research design, followed by descriptive or experimental research. For instance, hypotheses generated through exploratory research can be tested using descriptive or experimental methods.
Not every research project needs to begin with exploratory research. The starting point depends on how well the problem is defined and how confident the researcher is in their approach. For example, a regularly conducted consumer satisfaction survey might directly start with descriptive or experimental research without needing an exploratory phase.
While exploratory research is typically the first step, this isn't always the case. It can also be used after descriptive or experimental research. For example, if descriptive or exploratory studies produce results that are difficult for managers to interpret, exploratory research can provide additional insights to clarify those findings.

4.4.3 Cross Sectional Research Design

As stated earlier, descriptive research is further subdivided into two categories: cross-sectional studies and longitudinal studies. The cross-sectional study examines a particular segment of the population that is being studied. It entails the collection of information from a specific sample of

population elements only once. They may be either single cross-sectional or multiple cross-sectional. In single cross-sectional designs, a single sample of respondents is selected from the target population, and information is obtained from this sample only once. These designs are also referred to as sample survey research designs. In multiple cross-sectional designs, information is obtained from each sample only once, and there are two or more samples of respondents. Information is frequently acquired at varying intervals from various samples.

There are two essential characteristics of cross-sectional studies:

- The cross-sectional study is carried out at a single moment in time, and thus applicability is most relevant for a specific period. For example, a cross-sectional study on the attitude of Indians towards hygiene pre and during CoronaVirus Disease (COVID) was vastly different, and a study done in 2020 revealed a different attitude and behaviour towards hygiene, which might not be absolutely in line with that found earlier.

- Secondly, these studies are conducted on a specific subset of respondents from the population unit under investigation (e.g., loyal customers of a product, employees of an organization). This sample is currently being evaluated and investigated solely for the purpose of the study's time coordinate.

A cross-sectional survey, which is conducted on sample groups at different time intervals, is called cohort analysis. A cohort is a group of respondents who experience the same event within the same time interval. For example, a birth (or age) cohort is a group of people who were born during the same time interval, such as 1981–1996 for Millennials or 1997–2012 for Gen Z. The term cohort analysis refers to any study in which there are measures of some characteristics of one or more cohorts at two or more points in time. For example, in the above case of the attitude of Indians towards hygiene, if we study and compare the attitudes of Millennials versus Gen Z towards hygiene during COVID, it would be a cohort analysis.

Cross-sectional studies are advantageous for investigating present patterns of behaviour or opinion. Nevertheless, it is not a prudent decision to consider the respondent's likelihood of making future decisions or to delve too deeply into the past in order to distinguish between the present and past behavior. In such instances, it is preferable to implement a study that is designed to gather information at various points in time. The results would be more valid and reliable. The benefit would be that, rather than relying on the respondent's memory or prediction, an actual surveillance of behaviour patterns would occur over time.

4.4.4 Longitudinal Research Design

A longitudinal study design is appropriate when a single sample of a population is observed over an extended period. For example, a panel of selected consumers might be monitored over time to study their grocery buying habits. In other words, a fixed sample (or samples) of population elements is measured repeatedly in longitudinal research design.

Key features of longitudinal studies include:

- The selection of a representative group that reflects the population being studied.
- Repeated measurements of the group at regular intervals, focussing on specific variables of interest.
- Once the sample is selected, it should remain consistent throughout the study. If a panel member drops out, they should be replaced by someone who similarly represents the population to maintain the integrity of the study.

In contrast to a cross-sectional design, a longitudinal design maintains the same sample or samples across time. In other words, the same subjects are investigated again. Longitudinal research provides a series of "pictures," as opposed to the standard cross-sectional design, which provides a snapshot of the variables of interest at a particular moment in time. These "pictures" provide a comprehensive overview of the circumstances and the changes that occur throughout time. Though longitudinal and cross-sectional research are visualised conceptually as two ends of a continuum, in practice, the two might merge or complement each other in usage.

For example, imagine a business school has launched a new postgraduate program in human resource management. The school wants to evaluate how different stakeholders—students, recruiters, and faculty—perceive the program's structure and the quality of students. They also aim to monitor and adjust the program based on changes in these perceptions over time. For instance, the school could assess these attitudes every six months: once during placements and again six months after the graduates start working. For such an objective, a longitudinal study would be most suitable. The recruiter population, for example, could be consistently studied over time. However, when it comes to evaluating student outcomes, a cross-sectional study might be more practical. Each year, the school would need to assess the perspectives of that specific graduating batch rather than maintaining a fixed panel of students. On the other hand, faculty members could form a stable panel, allowing the school to monitor changes in their opinions over time.

The difference between longitudinal research and cross-sectional research can be understood from Table 4.5.

In summary, for tracking changes in measured variables over time, a longitudinal design is ideal. Still, a mix of cross-sectional and longitudinal designs might be necessary depending on the stakeholder group being analyzed. The longitudinal research design is also referred to as the time-series design due to the repeated measurement over time.

Table 4.5: Difference between longitudinal research and cross-sectional research

Aspect	Longitudinal Research	Cross-Sectional Research
Definition	Studies the same group or variable over a long period of time.	Studies a group or variable at a single point in time.
Time Frame	Conducted over an extended period.	Conducted at one specific moment.
Purpose	Tracks changes or trends over time.	Provides a snapshot or overview of a situation.
Data Collection	Data is collected multiple times from the same sample.	Data is collected only once from the sample.
Examples	Measuring the spending habits of a group over 5 years.	Surveying people's opinions about a product today.
Cost and Time	Usually more time-consuming and costly.	Less time-consuming and generally more affordable.
Suitability	Ideal for studying developments, patterns, or effects over time.	Ideal for obtaining quick, descriptive information.
Sample	Involves repeated interaction with the same participants.	May involve different participants within the target group.

Check Your Progress B:

1) Explain the characteristics of cross-sectional research design.

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

2) Explain extraneous variables.

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

3) What is descriptive research design? Explain its purpose.

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

4) State whether the following statements are true (T) or false (F).

   a) Research designs are the blueprint of the research study to be conducted.

   b) The definition of variables and constructs under investigation can be facilitated by exploratory research designs.

   c) A study to analyse the profile of the supporters of political parties would need a cross-sectional research design.

   d) A longitudinal design is distinguished by the testing of various groups of individuals over a singular period of time.

   e) In a longitudinal research design, the research variable is examined at predetermined intervals of time.

   f) Descriptive designs do not require any quantitative statistical analysis.

   g) Time series analyses a form of longitudinal design

4.5 LET US SUM UP

This unit discusses the importance of research design in determining the research problem and choosing the most effective methodology for achieving research goals. There are three main categories of research design: exploratory, conclusive, and experimental. Research design is a comprehensive blueprint outlining specific strategies and methods to accomplish research objectives. It serves as the foundation for conducting a study, guiding the research process, and providing a framework for data collection and analysis. A well-planned research design ensures the dependability and accuracy of results, ensuring consistency and stability of findings. A robust research design reduces mistakes, biases, and inconsistencies, ensuring trustworthy results and aligning the study with real-world scenarios. In summary, research design is crucial for a successful research project and ensuring the accuracy and relevance of results.

An effective research design is crucial for achieving meaningful, actionable, and reliable results in business research. It involves a harmonious theoretical and conceptual framework with the research aims, and aligns with data collection methods. Thorough planning is essential to minimise errors and ensure the project's success. Research design helps in reducing inaccuracies, improving efficiency, eliminating biases, minimising time wastage, facilitating material collection, testing hypotheses effectively, resource planning, expert review, and guidance in the right direction.

There are several types of research designs, including exploratory and conclusive designs. Exploratory research is the simplest and most flexible research method, used when the research topic is difficult to measure or when standard measurement methods don't fully capture certain aspects. It is useful for understanding the unknown, problem definition and hypothesis formation, concept identification and exploration, preliminary screening in product development, understanding belief and attitude structures, examining

underlying reasons for statistical differences, addressing sensitive issues, and probing deeply held beliefs or emotions.

Conclusive research design is a quantitative method used to provide definitive insights for decision-making or drawing conclusions. It involves extensive, representative samples and quantitative analysis of collected data. It can be classified as descriptive or experimental and can be further classified as cross-sectional or longitudinal. Descriptive research design aims to describe a phenomenon under study, focussing on the population of individuals being examined. Examples of descriptive studies in business research include market analysis, market share analysis, sales analysis, product usage analysis, and advertising research.

Experimental research design is used to infer a causal relationship between a phenomenon and its consequences. It is suitable for establishing causal relationships, ascertaining the nature of the link between causal factors and the expected results, and testing hypotheses. Experimental research design involves manipulating independent variables, test units, dependent variables, and extraneous variables to measure their effects on dependent variables. It requires a structured approach similar to descriptive research,but is not suitable for examining causal relationships. Experimentation is established when the researcher manipulates independent variables and evaluates their impact on dependent variables while controlling extraneous variables.

Cross-sectional and longitudinal studies are two types of descriptive research. Cross-sectional studies examine a specific segment of the population and collect information from a single sample of respondents. They are useful for investigating present behavioural or opinion patterns but may not be suitable for examining future decisions or past behaviour. Longitudinal studies observe a single sample of a population over an extended period, maintaining the same sample or samples over time. Key features of longitudinal studies include selecting a representative group, repeating measurements at regular intervals, and maintaining consistency in the sample. For tracking changes in measured variables over time, a longitudinal design is ideal. However, a mix of cross-sectional and longitudinal designs may be necessary, depending on the stakeholder group being analysed.

4.6 KEYWORDS

Exploratory research design is used in instances where the subject of the study cannot be measured in a quantitative manner or where the process of measurement cannot realistically represent particular qualities.

Conclusive research design aims to provide a detailed description of certain events, to test specific hypotheses, and to thoroughly analyze specific relationships.

Descriptive research design is a type of conclusive research design that aims to describe a phenomenon under study.

Experimental research design is a conclusive research design that is primarily designed to gather evidence regarding the causal relationship between a phenomenon and its consequences.

Cross-sectional research design is a type of descriptive research design that examines a particular segment of the population that is being studied.

Longitudinal research design is a type of descriptive research design apt when a single sample of a population is observed over an extended period.

4.7 ANSWERS TO CHECK YOUR PROGRESS

Answer to check your progress (B)

a) True b) True c) True d) False e) True f) False g) True

4.8 TERMINAL QUESTIONS

1. How would you describe a research design? What are the key components of a research design? Provide examples to support your explanation.
2. 'Majority of the research designs are exploratory cum descriptive in nature in business research.' How?
3. What are the primary objectives of exploratory research?
4. Differentiate between longitudinal and cross-sectional research designs. In which circumstances would you suggest that one be used over the other?
5. What is experimental research design? Define its purpose.
6. Differentiate between exploratory and conclusive research design.
7. Explain the relationship between exploratory, descriptive, and experimental research design.

4.9 FURTHER READINGS

Myers, J. L., Well, A. D., & Lorch Jr, R. F. (2013). *Research design and statistical analysis*. Routledge.

Kothari, C. R. (2004). Research methodology.

Malhotra, N. K. (2020). Marketing research: an applied prientation. pearson.

Dannels, S. A. (2018). Research design. In *The reviewer's guide to quantitative methods in the social sciences* (pp. 402-416). Routledge.

Note	These questions are helpful to understand this unit. Make an effort to write the answer to these questions but do not send your answer to university. It is only for your practice.

BLOCK 2
DATA COLLECTION AND SAMPLING

BLOCK 2 DATA COLLECTION AND SAMPLING

This block provides learners with a comprehensive understanding of data collection methods, measurement techniques, questionnaires, and sampling. It explores the nature of data, sources, advantages and disadvantages of various data types, emerging data sources, and measurement scales. The block also covers the design and significance of questionnaires, as well as the fundamentals of sampling, including methods, population and sample characteristics, sampling frame, unit, sample size, and the central limit theorem. These topics equip students with the skills needed to make informed decisions for their research studies.

The block, titled **"Data Collection and Sampling,"** consists of four units, as detailed below:

Unit 5: Collection of Data

This unit introduces the basics of data, including its definition, collection methods, and role in research. Learners will gain an understanding of the significance of data collection, the nature of data, and its various sources. The unit also discusses the advantages and disadvantages of each type of data source, as well as emerging data sources in research.

Unit 6: Measurement and Scaling Techniques

This unit focuses on the techniques used to measure and scale data in research. It explains different levels of measurement scales, such as nominal, ordinal, interval, and ratio scales, as well as comparative and non-comparative scaling techniques. Learners will also gain insights into the concepts of reliability and validity, enabling them to evaluate and select the most appropriate scaling techniques for their research studies.

Unit 7: Questionnaire Design

This unit delves into the importance of questionnaires in research. It covers the meaning and significance of questionnaires, along with the steps involved in designing and constructing them. Learners will also explore the key features of an effective questionnaire and understand the role of pre-testing and pilot testing in refining questionnaire designs for research.

Unit 8: Sampling Design

This unit introduces learners to the concept of sampling and its importance in research. It covers the relationship between a population and a sample, the characteristics of a good sample, and the concept of a sampling frame and sampling unit. Additionally, the unit explains various sampling methods, determining sample size, and the central limit theorem, providing learners with the tools needed to develop robust sampling strategies for their research.

UNIT 5 COLLECTION OF DATA

Structure

5.0 OBJECTIVES

After studying this unit, you will be able to:

- Understand the importance of data collection in business research;
- Differentiate between qualitative and quantitative data;
- Explain the distinction between primary and secondary data;

- Identify the various methods of collecting primary data and their applications;
- Analyse the merits and limitations of secondary data; and
- Explore emerging data sources and their significance in modern research.

5.1 INTRODUCTION

In the preceding units, we explored how to identify a research problem and formulate a research design. With the research design in place, the next critical step is the collection of data, which serves as the backbone of any research. Data collection involves systematically gathering, measuring, and analysing information to support research objectives, validate findings, and ultimately draw meaningful conclusions. Accurate and relevant data form the foundation for informed decision-making, enhancing the credibility and reliability of the research outcomes.

For instance, consider a company planning to launch a new mobile app. To ensure its success, the company must collect data about user preferences, emerging technology trends, and competitive offerings. Such information enables them to understand the market landscape, identify gaps, and design a product that aligns with user needs and expectations. Without reliable data, even the most well-thought-out strategies may fail to achieve desired outcomes.

Data collection methods vary widely, depending on research objectives, the nature of the study, and available resources. These methods can include direct interactions with individuals through tools like surveys, interviews, and focus groups, or the utilization of secondary data from sources such as government reports, company records, or online databases. While each method offers unique advantages, they also come with challenges such as bias, cost, and time considerations. Selecting the most appropriate method requires a careful assessment of research goals and practical constraints.

This unit delves into the nature, sources, and methods of data collection, providing a comprehensive overview of how data is obtained for research. Additionally, it highlights the strengths and limitations of various methods, offering valuable insights into selecting the most suitable approach for specific research contexts. By the end of this unit, learners will be equipped to design effective data collection strategies that ensure the integrity and validity of their research.

5.2 SIGNIFICANCE OF DATA COLLECTION

Collecting data is a vital step in research, as it forms the foundation for understanding problems, finding solutions, and drawing accurate conclusions. Properly collected data ensures that decisions and insights are based on factual evidence rather than assumptions, making it essential in any systematic study or business decision-making process.

Data collection provides facts and evidence that validate research findings and support informed conclusions. For instance, if a researcher aims to determine whether students prefer online learning, gathering data such as survey responses and attendance records offers objective insights into their preferences and behaviour. This not only strengthens the credibility of the research but also helps in identifying actionable solutions.

Additionally, data collection plays a crucial role in decision-making. Businesses, for example, often rely on data to guide their strategies. Before opening a new store, a company may analyse data on local population demographics, purchasing power, and shopping habits to ensure the decision aligns with market demand. By basing decisions on reliable data, organisations reduce the risks associated with uncertainty and improve their chances of success.

Data also helps identify patterns and trends, enabling researchers and businesses to understand what is happening and predict future developments. *For example, a store analysing its sales data might observe that winter clothes sell more in November, allowing them to stock up and prepare for the seasonal demand. Such insights are essential for optimizing operations and addressing customer needs effectively.*

Finally, carefully collected data builds trust in research findings. Accurate and unbiased information lends credibility to the conclusions drawn, making them more reliable. For instance, a teacher conducting a survey to gather students' opinions on their favourite subjects can achieve better results and make more informed decisions than relying on assumptions or informal observations.

In summary, data collection ensures that research and decision-making processes are grounded in facts rather than guesses, ultimately contributing to more effective problem-solving, strategic planning, and actionable outcomes.

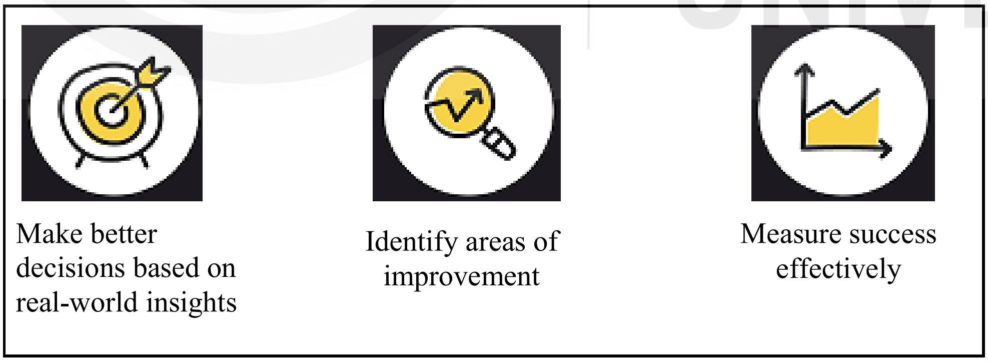


Figure 5.1: Importance of Data collection

5.3 NATURE OF DATA

Data serves as the cornerstone of any research, enabling analysis, interpretation, and meaningful conclusions. Depending on its characteristics and purpose, data can be classified into two main types: **Qualitative** and **Quantitative**. Each type has distinct features, sources, and applications in research and decision-making.

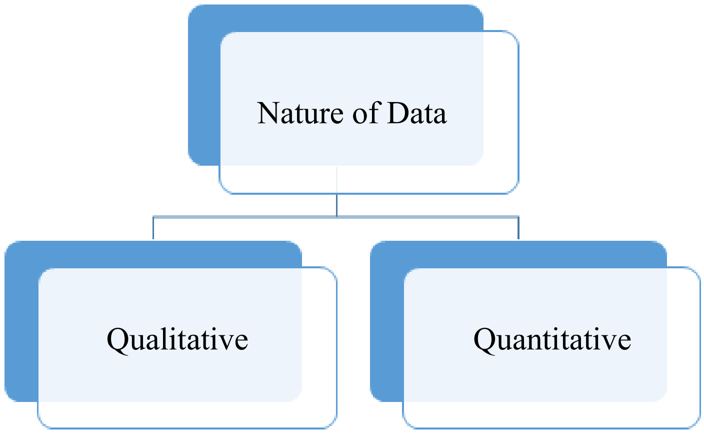


Figure 5.2: Nature of Data

5.3.1 Qualitative Data

Qualitative data refers to non-numerical information that captures the essence of opinions, experiences, and behaviours. It helps researchers understand "why" or "how" people think or act in a particular way, focusing on depth and context.

Example: *If students are interviewed about their experiences with online shopping, responses such as "I find online shopping flexible because they save travel time" are examples of qualitative data. Such data offers rich insights into their feelings and perceptions.*

Common Sources: Interviews, focus group discussions, and open-ended survey questions. We will study these in detail further in this unit.

5.3.2 Quantitative Data

Quantitative data involves numbers or measurable facts that can be analysed statistically. It is structured, objective, and focuses on "how much" or "how many." Quantitative data is ideal for identifying patterns, testing hypotheses, and making comparisons.

Example: *Recording students' grades in online and offline exams provides quantitative data. A result like "80% of students scored above 75 marks" allows researchers to evaluate performance trends statistically.*

Common Sources: Surveys with numerical options, test scores, and government statistics.

Table 5.1 Key Differences between Qualitative Data and Quantitative Data

Aspect	**Qualitative Data**	**Quantitative Data**
Nature	Descriptive	Numerical
Purpose	To explore and understand "why" or "how"	To measure "how much" or "how many"
Examples	Customer feedback, interview transcripts	Sales figures, exam scores
Analysis	Looking for themes, understanding text data	Statistical analysis, charts, graphs
Data Format	Text, audio, or visual formats	Numbers, percentages, or ratios

Both types of data often work together in research to provide a holistic view of a problem. *For example, a company might use qualitative data to understand customer emotions about a product and quantitative data to measure satisfaction levels across a broad audience. This combination strengthens the depth and accuracy of research findings.*

In summary, while qualitative data provides context and depth, quantitative data offers precision and scalability. Understanding the nature of both is essential for effective research design and implementation.

5.4 SOURCES OF DATA

Data can be collected from various sources depending on the purpose and scope of the research. These sources are broadly classified into two categories: **Primary Data Sources** and **Secondary Data Sources.** Each source has unique features and uses, and the choice depends on the research objectives, time constraints, and resources available.

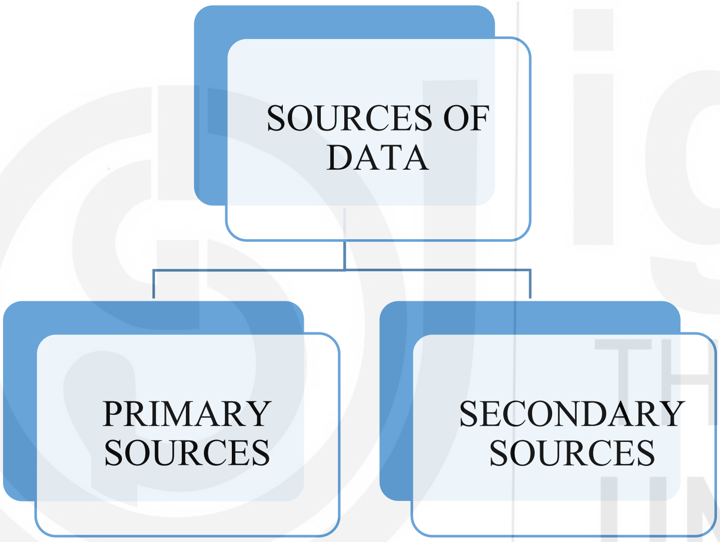


Figure 5.3: Sources of Data

5.4.1 Primary Sources

Primary data refers to original information collected firsthand by the researcher specifically for the purpose of the study. It is tailored to address specific research questions and is generally considered more accurate and relevant to the objectives of the research.

Example*: Conducting a survey to understand how students use social media for learning.*

Common Methods: Common methods of primary data collection include:

1. Observations
2. Interviews
3. Surveys
4. Questionnaires

5.4.2 Secondary Sources

Secondary data refers to information that has already been collected, processed, and made available by other individuals or organisations. This data is reused to support new research objectives. Secondary data is often easier and faster to access, but researchers must ensure its relevance, accuracy, and credibility.

***Example**: Using a government report on unemployment to study job trends in urban areas is an example of secondary data, as the information has already been collected and published.*

Common Sources:

1. Books and journals
2. Government publications
3. Online databases and websites

Choosing the Right Source

Researchers must carefully decide whether to use primary or secondary data based on the specific objectives, scope, and constraints of their study. Both types of data have their unique advantages and limitations, and the choice often depends on the nature of the research question, the availability of resources, and the time frame for completing the study. In many cases, researchers combine both primary and secondary data to ensure a more comprehensive and well-rounded analysis.

For example, a company may conduct a customer survey (primary data) to gather specific insights about preferences for a new product while simultaneously consulting industry reports (secondary data) to analyse broader market trends. This dual approach ensures that the study benefits from both detailed firsthand information and broad contextual knowledge.

When to Use Primary Data:

Primary data is the ideal choice in situations where tailored, specific, or firsthand information is required to address unique research questions. It allows researchers to design the data collection process to meet the exact needs of the study.

***Example**: Conducting a focus group with users to gather feedback on the functionality and user experience of a newly launched mobile app.*

Other Use Cases:

1. When studying emerging issues where no existing data is available.
2. When the research requires direct interaction with respondents to collect real-time, accurate insights.
3. When researchers need detailed data on specific behaviours, preferences, or attitudes.

When to Use Secondary Data:

Secondary data is most suitable when existing information can fulfil the research objectives efficiently. It is particularly useful for broad or exploratory studies, where gathering new data might be costly, time-consuming, or unnecessary.

***Example**: Using government statistics to analyse employment trends across various regions instead of conducting an extensive survey.*

Other Use Cases:

1. For historical or longitudinal studies, where past records provide valuable insights.
2. When studying large-scale phenomena, such as demographic patterns, using data from government reports or global organizations.
3. When researchers are constrained by time or budget and need quick access to reliable data.

Using both primary and secondary data enhances the research's depth and reliability by balancing specific insights with broader contextual understanding.

***Example**: A retail company might survey customers (primary data) to understand satisfaction levels while using secondary data from industry reports to benchmark these findings against competitors.*

Check Your Progress A:

1) What is the significance of data collection in research?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

2) Mainly there are two sources of data, comment.

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

3) Differentiate between qualitative and quantitative data.

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

4) How to decide the source of a data?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

5.5 SOURCES OF COLLECTING PRIMARY DATA

Primary data refers to information that is directly collected by a researcher for a specific purpose or study. Unlike secondary data, which is pre-existing and collected by others, primary data is tailored to meet the unique objectives of a research project. It is original, first-hand data that offers high relevance and accuracy for addressing specific research questions.

For instance, consider a retail company planning to expand into a new market. To understand the preferences and buying behaviour of local consumers, the company conducts surveys, organizes focus groups, and observes shopping patterns in the target area. This directly collected information is primary data, as it is customized to provide insights into the company's specific objectives. Primary data helps researchers address unique problems, test hypotheses, and gain actionable insights that are not readily available through secondary sources.

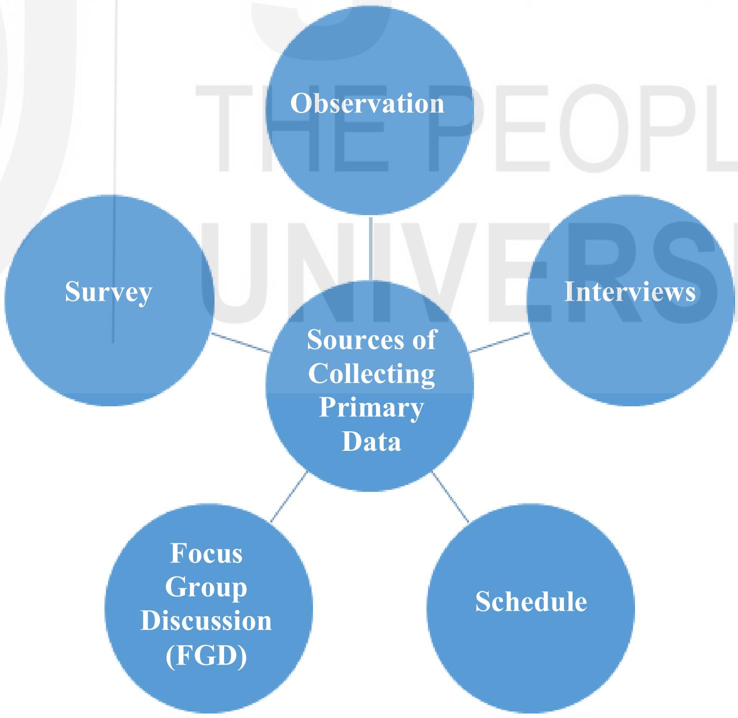


Figure 5.4: Sources of Primary Data

5.5.1 Survey

A survey is a method of data collection used to gather information from a large group of people through standardized tools, such as questionnaires or structured interviews. It is one of the most commonly used techniques in research, especially for studies requiring insights from a broad audience.

Surveys are versatile and can be conducted in-person, over the phone, online, or via mail, depending on the research objectives and resources available. *For example, A company surveys its customers to understand preferences for eco-friendly products.*

Key Features of Surveys

- **Large-Scale Data Collection**: Surveys are ideal for gathering data from a wide demographic or geographic area.
- **Standardized Tools**: Questions are pre-designed and consistent for all respondents, ensuring uniformity in data collection.
- **Diverse Formats**: Surveys can be conducted through various mediums, such as face-to-face interactions, online platforms, or paper-based forms.

Example*: A company conducting a survey to understand public opinion on electric vehicles can distribute an online questionnaire asking participants about their awareness, willingness to adopt, and preferences for electric vehicles. This approach allows the company to collect responses from a large group quickly and analyse trends effectively.*

Merits:

1. **Wide Coverage**: Surveys can reach a large population, making them ideal for generalizing results to a broader audience.

   Example*: A national survey on consumer preferences for digital payment systems can provide insights into trends across different regions.*
2. **Cost-Effective for Online Surveys**: Online surveys reduce logistical costs and simplify data collection.
3. **Standardization**: The uniformity of questions ensures consistency in the responses, facilitating easy comparison and statistical analysis.
4. **Flexibility**: Surveys can be adapted to different formats (e.g., online, telephone, or face-to-face), depending on the target group and available resources.

Limitations:

1. **High Costs for Large-Scale Studies**: Surveys requiring personal visits, such as those in remote areas, can be expensive.

   Example*: Conducting face-to-face surveys for rural development studies often involves significant travel and logistical expenses.*
2. **Low Response Rates**: Online or mail surveys may experience poor participation unless respondents are incentivized.
3. **Superficial Responses**: Participants may provide incomplete or inaccurate answers, especially if the survey is lengthy or unclear.
4. **Limited Depth**: Surveys are less effective for exploring complex issues or in-depth behavioural insights, as responses are typically constrained to predefined options.

5.5.1.1 Questionnaire

Under this method, questionnaires are sent personally or by post to various respondents with a request to answer the questions and return the questionnaire. If a questionnaire is sent by mail, it is called a **Mail Questionnaire**. Sometimes questionnaires may also sent via E-mail depending upon the nature of the study and availability of time and resources. After receiving the questionnaires, the respondents read the questions and record their responses in the space provided on the questionnaire. It is desirable to send the questionnaire with a self-addressed envelopes for quick and high rate of response.

Merits

1) You can use this method in cases where respondents are spread over a vast geographical area.
2) Respondents can take their own time to answer the questions. So, the researcher can obtain original data by this method.
3) This is an economical method because its mailing cost is less than the cost of personal visits.
4) This method is free from bias of the investigator as the information is given by the respondents themselves.
5) Large samples can be covered and thus the results can be more reliable and dependable.

Limitations

1) Respondents may not return filled in questionnaires, or they can delay in replying to the questionnaires.
2) This method is useful only when the respondents are educated and co-operative.
3) Once the questionnaire has been dispatched, the investigator cannot modify the questionnaire.
4) It cannot be ensured whether the respondents are truly representative.

5.5.1.2 Schedule

A Schedule is also a list of questions, which is used to collect the data from the field. This is generally filled in by the researcher or the enumerators. If the scope of the study is wide, then the researcher appoints people who are called enumerators for the purpose of collecting the data. The enumerators go to the informants, ask them the questions from the schedule in the order they are listed and record the responses in the space meant for the answers in the schedule itself. *For example, the population census all over the world is conducted through this method.* The difference between questionnaire and schedule is that the former is filled in by the informants, the latter is filled in by the researcher or enumerator.

Merits

Following are the main merits of schedule:

1) It is a useful method in case the informants are illiterates.
2) The researcher can overcome the problem of non-response as the enumerators go personally to obtain the information.
3) It is very useful in extensive studies and can obtain more reliable data.

Limitations

Following are the major limitations of schedule:

1) It is a very expensive and time-consuming method as enumerators are paid persons and also have to be trained.
2) Since the enumerator is present, the respondents may not respond to some personal questions.
3) Reliability depends upon the sincerity and commitment in data collection.

5.5.1.3 Difference between Questionnaire and Schedule

The difference between a questionnaire and a schedule lies in how the data is collected and who fills in the responses. While both are tools for gathering information, their applicability depends on the literacy level of respondents, the complexity of the study, and the researcher's involvement in the data collection process.

Table 5.2: Difference between Questionnaire and Schedule

Aspect	Questionnaire	Schedule
Definition	A set of questions filled out by the respondents themselves.	A list of questions filled out by the researcher or enumerator on behalf of the respondent.
Method of Collection	Sent via mail, email, or online platforms to respondents.	Researchers or enumerators personally visit respondents to record answers.
Respondents' Role	Respondents read and answer the questions independently.	Respondents answer verbally, and the enumerator records the responses.
Cost	Cost-effective, especially for large samples or geographically spread respondents.	Expensive due to the need for trained enumerators and travel expenses.
Response Rate	May have a lower response rate due to lack of follow-up.	High response rate as the enumerator ensures participation.
Suitability	Suitable for literate and cooperative respondents.	Suitable for illiterate or less cooperative respondents.
Examples	Feedback forms sent to customers about a product or service.	Researchers visiting homes to collect data for local business projects.

5.5.2 Interviews

An Interview is one of the most powerful tools and most widely used method for primary data collection in business research. In our daily routine we see interviews on T.V. channels pod cast & You Tube channels on various topics related to social, business, sports, budget etc. In the words of C. William Emory, 'personal interviewing is a two-way purposeful conversation initiated by an interviewer to obtain information that is relevant to some research purpose'. Thus, an interview is basically, a meeting between two persons to obtain the information related to the proposed study. The person who is interviewing is named as interviewer and the person who is being interviewed is named as informant. It is to be noted that the research data/information collected through this method is not a simple conversation between the investigator and the informant, but also the glances, gestures, facial expressions, level of speech etc., are all part of the process. Through this method, the researcher can collect varied types of data intensively and extensively.

Interviews can be classified as direct personal interviews and indirect personal interviews. Under the techniques of **direct personal interview**, the investigator meets the informants (who come under the study) personally, asks them questions pertaining to enquiry and collects the desired information. Thus, if a researcher intends to collect the data on spending habits of Indira Gandhi National Open University (IGNOU) students, he/ she would go to the Regional Center (RCs) of IGNOU to contact the students, interview them and collect the required information.

Indirect personal interview is another technique of interview method where it is not possible to collect data directly from the informants who come under the study. Under this method, the investigator contacts third parties or witnesses, who are closely associated with the persons/situations under study and are capable of providing necessary information. *For example, an investigation regarding a bribery pattern in an office.* In such a case it is inevitable to get the desired information indirectly from other people who may be knowing them.

Similarly, clues about the crimes are gathered by the Central Bureau of Investigation (CBI). Utmost care must be exercised that these persons who are being questioned are fully aware of the facts of the problem under study, and are not motivated to give a twist to the facts.

Another technique for data collection through this method can be structured and unstructured interviewing. In the **Structured interview** set questions are asked and the responses are recorded in a standardised form. This is useful in large scale interviews where a number of investigators are assigned the job of interviewing. The researcher can minimise the bias of the interviewer. This technique is also named as formal interview. In **Unstructured interview**, the the investigator may not have a set of questions but have only a number of key points around which to build the interview. Normally, such types of interviews are conducted in the case of an explorative survey where the researcher is not completely sure about the type of data he/she collects. It is also named as informal interview. Generally, this method is used as a

supplementary method of data collection in conducting research in business areas.

Nowadays, telephone or cell phone interviews are widely used to obtain the desired information for small surveys. For instance, interviewing credit card holders by banks about the level of services they are receiving. This technique is used in industrial surveys especially in developed regions.

Merits

The major merits of this method are as follows:

1) People are more willing to supply information if approached directly. Therefore, personal interviews tend to yield high response rates.
2) This method enables the interviewer to clarify any doubt that the interviewee might have while asking him/her questions. Therefore, interviews are helpful in getting reliable and valid responses.
3) The informant's reactions to questions can be properly studied.
4) The researcher can use the language of communication according to the standard of the information, so as to obtain personal information of informants which are helpful in interpreting the results.

Limitations

The limitations of this method are as follows:

1) The chance of the subjective factors or the views of the investigator may come in either consciously or unconsciously.
2) The interviewers must be properly trained, otherwise the entire work may be spoiled.
3) It is a relatively expensive and time-consuming method of data collection especially when the number of persons to be interviewed is large and they are spread over a wide area.
4) It cannot be used when the field of enquiry is large (large sample).

Precautions: While using this method, the following precautions should be taken:

1) Obtain thorough details of the theoretical aspects of the research problem.
2) Identify who is to be interviewed.
3) The questions should be simple, clear and limited in number.
4) The investigator should be sincere, efficient and polite while collecting data.
5) The investigator should be of the same area (field of study, district, state etc.).

5.5.2.1 Focus Group Discussion

A focus group discussion involves gathering individuals with similar backgrounds or experiences to have an in-depth conversation about a

particular subject. You might think of it as a kind of qualitative research, in which you inquire about people's thoughts and opinions.

Participants in a focused group discussion (FGD) are encouraged to engage in conversation with other group members, unlike in other research methodologies. In most cases, the organiser of the FGD interviews a small group of 8 to 12 persons in this manner. A moderator or interviewer discusses various subjects of interest in a non-linear fashion. Refer to Figure 5.5 for Focus Group Discussion.

Figure 5.5: Focus Group Discussion

Below are the main features of focused group discussions (FGDs).

- It involves an organised conversation with a group of people to learn about their ideas and experiences on a particular subject matter.
- An FGD is suitable for collecting several viewpoints on the same subject.
- This method makes it possible to get insights into people's common knowledge of daily life and how individuals are affected when they are in groups.
- Group leadership and interpersonal skills are critical to the success of a moderator.

Merits

- Discussions that are open and free generate fresh ideas that may be quite helpful in making decisions.
- Focus groups are dynamic. The moderator may adjust the group conversation to make it easier for everyone to participate. With this dynamic, a focus group's findings are more accurate.
- Gestures and energising activities may help researchers get new insights into their subjects.

Limitations

- They are more expensive because the organiser has to compensate the participants.
- Participants feel like they are under supervision, so they do not put their honest opinion in front.
- The moderator may be disappointed because the results may be biased.

5.7 ADVANTAGES AND DISADVANTAGES OF PRIMARY DATA

Primary data offers researchers the ability to collect highly specific and relevant information tailored to their study's objectives. However, it comes with challenges such as higher costs and time requirements. Below is an overview of its advantages and disadvantages:

Advantages of Primary Data

1. **Specific to the Research Purpose**:

   Primary data is collected with a specific goal, making it highly relevant and tailored to the study.

   Example: *Conducting a survey to know students' preferences for online or offline classes ensures the data directly relates to the research question.*

2. **Accurate and Reliable**:

   Since the researcher collects the data firsthand, it is generally accurate and free from manipulation.

   Example: *Observing customer* behaviour in a store gives first hand insights into shopping patterns.

3. **Up-to-Date Information**:

   Primary data reflects current conditions, making it more reliable for recent trends.

   Example: *Conducting interviews with farmers about current challenges in agriculture ensures updated insights.*

4. **Greater Control**:

   The researcher can control how the data is collected, ensuring it meets the study's requirements.

   Example: *A teacher interviews students about their learning styles to design better classes.*

Disadvantages of Primary Data

1. **Time-Consuming**:

   Collecting primary data requires significant time for planning, conducting, and analysing.

   Example: *Conducting a nationwide survey on healthcare facilities may take months to complete.*

2. **Expensive**:

   Primary data collection involves costs for surveys, travel, and hiring enumerators.

   Example: *A market research company conducting in-person interviews incurs costs for transportation and personnel.*

3. **Limited Coverage**:

   It may not be feasible to collect primary data from large populations due to resource constraints.

   Example: *A small company might struggle to survey all its customers due to budget limitations.*

4. **Possibility of Bias**:

   Researchers' personal biases or improper techniques can affect data quality.

   Example: *An interviewer leading the respondents to specific answers might result in biased data.*

5.8 SOURCES OF COLLECTING SECONDARY DATA

Secondary data refers to information that has already been collected, processed, and published by others, making it a valuable resource for research that saves time and reduces costs. This type of data is particularly useful for exploring broad trends, forming preliminary insights, or complementing primary research efforts. Secondary data can be accessed from a variety of sources, which can be broadly classified into internal and external sources as shown in Figure 5.4 and below sub-section.

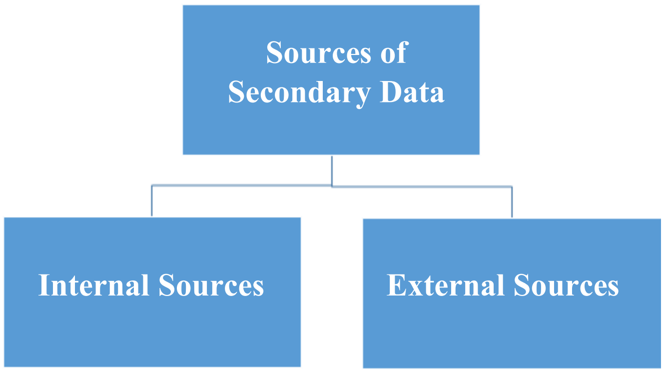


Figure 5.4: Sources of Collecting Secondary Data

5.8.1 INTERNAL SOURCES

These are data collected within an organization or institution, Examples:

1. Sales records
2. Employee performance reports
3. Financial statements
4. Customer feedback collected over time

Use Case: A company can analyse its past sales data to predict future demand for its products.

5.8.2 EXTERNAL SOURCES

These are data obtained from outside the organization or institution, Examples:

1. **Government Publications:** Census reports, economic surveys, and trade statistics.

   Example: Using census data to study population growth in urban areas.

2. **Research Journals:** Published research papers and academic studies.

   Example: Referencing journal articles to understand market trends.

3. **Websites and Databases:** Online resources like government websites, data repositories, or global organizations.

   Example: Using WHO reports for health-related studies.

4. **International Agencies:** Reports from organizations like the World Bank, IMF, or United Nations.

   Example: Using World Bank reports to analyse global poverty rates.

Key Points to Consider While Using Secondary Data

1. Ensure the data is relevant to your research objectives.
2. Check the reliability and credibility of the source.
3. Verify whether the data is up-to-date.

5.9 ADVANTAGES AND DISADVANTAGES OF SECONDARY DATA

There are certain advantages and disadvantages in using the secondary data. Let us now discuss them.

Advantages of Secondary data: Following are the main advantages of secondary data:

1. Cost-Effective

Secondary data is more economical as it eliminates the need for expenses like printing data collection forms or hiring large teams of enumerators.

2. **Time-Saving**

   Secondary data can be obtained quickly, often in just a few days, whereas primary data collection may take months to complete. This makes it ideal for time-sensitive investigations.

3. **Facilitates Large-Scale Studies**

   Secondary data is useful when collecting primary data is not feasible. For example, data on national income or census statistics, which are challenging for individual researchers to gather, are readily available through official publications.

4. **Access to Global Data**

   Secondary data provides access to worldwide information on diverse topics, such as trade, population, health, and industry, through sources like the United Nations, World Bank, or International Monetary Fund.

5. **Enables New Insights**

   Investigators can utilize pre-existing data to gain valuable insights into their research problems, often discovering patterns or relationships that might have been overlooked.

6. **Widely Used for Statistical Analysis**

   Many statistical analyses rely on secondary data due to its availability across various fields. Primary data is often used only when secondary data is insufficient for analysis.

Disadvantages of Secondary data: Following are the main disadvantages of secondary data:

1) **Risk of Reliability Issues**

   Secondary data may not always be reliable or suitable, so researchers must verify its accuracy. If these aspects are not verified, the results may be inaccurate or misleading.

2) **Mismatch with Research Needs**

   It can be challenging to find secondary data that precisely fits the specific requirements of a study.

3) **Accuracy Concerns**

   Secondary data may suffer from errors such as bias, an inadequate sample size, or poor definitions, which can affect its accuracy and reliability.

4) **Unavailability in Certain Cases**

   There are situations where secondary data is unavailable, forcing researchers to collect primary data despite its higher cost and time demands.

5.10 EMERGING DATA SOURCES

With the rapid advancement of technology, modern research has gained access to innovative data sources that simplify data collection and broaden its scope. These sources, such as datasets and repositories, offer pre-organized and electronically stored data, making research more efficient and comprehensive. Emerging data sources provide ready-to-use information for large-scale or global analysis, saving researchers time and resources while offering valuable insights for business and academic studies.

5.10.1 Dataset And Repositories

Datasets and repositories are organized collections of data, stored electronically, and made accessible for various research and analysis purposes. These sources are increasingly valuable in modern research due to their ease of access, breadth of information, and potential for data-driven insights. They provide researchers with pre-existing data, saving time and resources while enabling large-scale or global analysis.

Examples:

1. **Open-Source Databases:** Platforms like Kaggle and UCI Machine Learning Repository offer datasets for research and analysis.

   Example: A startup uses Kaggle to analyse customer preferences for mobile app features.

2. **Government Databases:** Many governments provide public access to data, such as census records or economic indicators.

   Example: India's National Sample Survey Office (NSSO) provides data on employment and household consumption.

3. **Global Organizations:** Agencies like the United Nations, IMF, and WHO offer international datasets on health, trade, and development.

   Example: WHO data on global vaccination rates can help analyse healthcare trends.

Benefits of Emerging Data Sources

1. **Accessibility**:

   Easily available online, often free or at a minimal cost.

   ***Example**: Open-access repositories allow researchers worldwide to access and use datasets without major expenses.*

2. **Diversity of Information**:

   Covers a wide range of topics and domains.

   ***Example**: Analysing Instagram data to study how students engage with career-related content.*

3. **Up-to-Date Information**:

   Emerging data sources often update in real-time, providing the most current insights.

***Example**: Websites tracking stock market trends update their datasets multiple times daily.*

4. **Support for Big Data Analytics**:

   Enables researchers to analyse massive datasets for better accuracy and insights.

   ***Example**: Analysing patterns in online shopping behaviour using millions of transaction records.*

Challenges of Emerging Data Sources

1. **Data Privacy**:

   Some datasets, such as those from social media, raise privacy concerns.

2. **Data Overload**:

   Researchers may struggle to filter relevant information from the vast amount of data available.

3. **Verification Issues**:

   The accuracy and authenticity of publicly available data must be carefully checked.

Check Your Progress B:

1. Answer the following questions:

1) Discuss the methods of collecting primary data. Compare the merits and demerits of observation and surveys as data collection tools.

   …………………………………………………………………………………

   …………………………………………………………………………………

   …………………………………………………………………………………

   …………………………………………………………………………………

2) How can researchers ensure the reliability of secondary data?

   …………………………………………………………………………………

   …………………………………………………………………………………

   …………………………………………………………………………………

   …………………………………………………………………………………

   …………………………………………………………………………………

3) Differentiate between a questionnaire and a schedule. Discuss their applications in business research.

   …………………………………………………………………………………

   …………………………………………………………………………………

   …………………………………………………………………………………

   …………………………………………………………………………………

4) What are datasets and repositories? How do they contribute to modern research, especially in the Indian context?

…………………………………………………………………………….

…………………………………………………………………………….

…………………………………………………………………………….

…………………………………………………………………………….

2. True or False

a) Primary data is collected firsthand by the researcher for a specific research purpose. (True/False)

b) Secondary data refers to information collected by someone else for a different purpose. (True/False)

c) Surveys are a primary data collection method that involves asking respondents a set of standardized questions. (True/False)

d) Observational studies are a secondary data collection method that involves watching subjects in their natural environment. (True/False)

e) Qualitative data is numerical and can be measured and quantified. (True/False)

3. Fill in the blanks

a) __________ data is collected directly from the source for a specific research purpose.

b) __________ data refers to information that has been previously collected for another purpose.

c) A __________ is a structured set of questions used to collect data from respondents.

d) __________ involves gathering data by observing subjects in their natural environment without interference.

e) __________ data is descriptive and provides detailed insights into a research subject**.**

5.11 LET US SUM UP

This unit emphasizes the critical role of data collection in business research, outlining its significance, sources, and methods. Data collection is the backbone of research, enabling systematic gathering, measurement, and analysis of information to support objectives, validate findings, and derive actionable insights. Properly collected data enhances decision-making, reduces uncertainty, and ensures credible outcomes.

The unit explores the nature of data, classifying it into qualitative and quantitative types. Qualitative data provides descriptive insights into behaviours and opinions, often collected through interviews and focus groups, while quantitative data offers measurable, numerical facts for statistical analysis. Both types are integral, providing depth and precision in research findings.

The sources of data are categorized into primary and secondary data. Primary data is original information collected first-hand through methods such as surveys, interviews, focus groups, observations, and schedules. It is tailored to specific research needs but is resource-intensive and time-consuming. Secondary data, on the other hand, is pre-existing information obtained from internal (e.g., sales records, financial reports) or external sources (e.g., government publications, research journals). While it is cost-effective and time-saving, researchers must ensure its relevance and reliability.

The unit also introduces emerging data sources, such as datasets and repositories, which provide pre-organized digital information for large-scale analysis. These sources are increasingly valuable in modern research, offering accessibility, diversity, and real-time updates.

Advantages and disadvantages of both primary and secondary data are discussed, helping researchers choose the most suitable approach based on objectives, scope, and constraints. By combining both types, researchers can achieve a comprehensive understanding of business problems.

In summary, this unit equips learners with the knowledge and skills to design effective data collection strategies, emphasizing the importance of accurate, relevant, and credible data in achieving meaningful research outcomes.

5.12 KEYWORDS

Data Collection: The process of gathering information for analysis and research purposes.

Dataset: A collection of data, often stored digitally, used for analysis.

Observation: A method of collecting data by watching and recording behaviours or events.

Primary Data: Original data collected firsthand for a specific research study.

Qualitative Data: Non-numerical data that focuses on descriptions and insights (e.g., opinions or behaviours).

Quantitative Data: Numerical data that can be measured and statistically analysed (e.g., percentages or scores).

Questionnaire: A written set of questions filled out by respondents to gather data.

Repositories: Online platforms or databases where datasets are stored and made accessible for research.

Secondary Data: Pre-existing data collected by others, used for new research.

Schedule: A set of questions filled by an enumerator during face-to-face interactions with respondents.

5.13 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress B

2. True/False
   a) True b) True c) True d) False e) False

3. Fill in the Blanks

   a) Primary b) Secondary c) Questionnaire d) Observation e) Qualitative

5.14 TERMINAL QUESTIONS

1. Define primary and secondary data. Provide examples to illustrate their differences.
2. Explain the significance of data collection in research and decision-making. Why is it considered a crucial step in any study?
3. Discuss the methods of collecting primary data. Compare the merits and demerits of observation and surveys as data collection tools.
4. What are the advantages and disadvantages of using secondary data? How can researchers ensure its reliability and relevance?
5. Describe qualitative and quantitative data. Provide examples to highlight their key differences and complementary roles in research.
6. What are datasets and repositories? How do they contribute to modern research, especially in the Indian context?
7. Explain the challenges associated with emerging data sources. How can researchers overcome these challenges while using Indian datasets?
8. Differentiate between a questionnaire and a schedule. Discuss their applications in business research.
9. When should researchers choose primary data over secondary data? Give examples of situations where a combination of both is beneficial.
10. Highlight the benefits of accessing government and global organization databases for research in India. Provide examples to support your answer.

5.15 FURTHER READINGS

1. Ajayi, V. O. (2017). Primary sources of data and secondary sources of data. *Benue State University*, *1*(1), 1-6.
2. Jackson, P. G. (1990). Sources of data. *Measurement issues in criminology*, 21-50.
3. Daneshjo, N., & Kravec, M. (2014). Sources of marketing information system. *International Journal of interdisciplinarity in theory and practice*, (5), 22-24.
4. Kothari, C. R. (2004). Research methodology.
5. Mittal, S., Chawla, D., & Sondhi, N. Journal of Indian Business Research.

Note	These questions are helpful to understand this unit. Make an effort to write the answer to these questions but do not send your answer to university. It is only for your practice.

UNIT 6 MEASUREMENT AND SCALING TECHNIQUES

Structure

6.0 OBJECTIVES

After studying this unit, you will be able to:

- Understand the concept of measurement and scaling;
- Know difference between nominal, ordinal, interval and ratio scale;
- Classify scaling techniques as comparative and non-comparative;

- Comprehend the criteria used for scale evaluation; and
- Understand reliability and validity of a scale

6.1 INTRODUCTION

Once the researcher has a clear understanding of what they wish to understand in their target respondents, they should consider the concepts of scaling and measurement. These concepts are vital for developing effective questionnaires or other instruments of measurement that will fulfil the research objectives in the most accurate manner. This unit describes the concepts of scaling and measurement and discusses the four primary scales of measurement: nominal, ordinal, interval and ratio. In this unit both comparative and non-comparative scaling techniques will also be discussed in detail. The comparative techniques, consisting of paired comparison, rank order, constant sum and Q-sort scaling, are discussed and illustrated with examples. The non-comparative techniques are composed of continuous and itemized rating scales. This unit also discusses and illustrates the popular itemized rating scales – the Likert, semantic differential and Stapel scales – as well as the construction of multi-item rating scales. We will also study how scaling techniques should be evaluated in terms of reliability and validity enabling the researcher to select a particular scaling technique.

In our daily lives, whenever we are said to measure, we use some yardstick to determine weight, height or some other feature of a physical object. We also measure when we judge how well we like a song, a painting or the personality of our friends. We thus measure physical objects as well as abstract concepts. Measurement is a relatively complex and demanding task, especially when it deals with qualitative or abstract phenomena like taste, honesty, intelligence, customer’s purchase intention, and brand loyalty etc. These characteristics are also called constructs.

It is easy to assign numbers in respect to properties of some objects, but it is relatively difficult with respect to abstract characteristics. For example, measuring someone's stress level is much more difficult than measuring their height or weight. It is because for measuring someone's height and weight we have some standard units of measurement but abstract construct like stress level cannot be measured directly.

6.2 SIGNIFICANCE OF MEASUREMENT AND SCALING

Before we proceed further, it will be worthwhile to understand the following two terms: (a) Measurement, and (b) Scaling.

a) **Measurement:** It means assigning numbers or other symbols to characteristics of objects according to certain pre-specified rules. It is important to note that we don’t measure the object but some characteristic of it. Likewise, in a survey we don’t measure the respondents but measure the respondent's characteristics like feelings, attitudes, opinions etc. For example, you may assign ‘1’ for Male and ‘2’

for Female respondents. Similarly in response to a question on whether he/she is using the ATM provided by a particular bank branch, the respondent may say ‘yes’ or ‘no’. You may wish to assign the number ‘1’ for the response ‘yes’ and ‘2’ for the response ‘no’. We assign numbers to these characteristics for two reasons. First, the numbers facilitate further statistical analysis of data obtained. Second, numbers facilitate the communication of measurement rules and results. The most important aspect of measurement is the specification of rules for assigning numbers to characteristics. The rules for assigning numbers should be standardized and applied uniformly. This must not change over time or objects.

b) **Scaling:** Scaling may be considered as an extension of measurement. It is the assignment of objects to numbers or semantics according to a rule. In scaling, the objects are text statements, usually statements of attitude, opinion, or feeling. For example, in a survey, a bank might ask customers how much they agree with the statement that the branch provides good service. Customers could respond with options like "strongly agree," "somewhat agree," "somewhat disagree," or "strongly disagree." To simplify analysis, these responses can be converted into numbers: "strongly agree" as 1, "somewhat agree" as 2, "somewhat disagree" as 3, and "strongly disagree" as 4. This way, each customer's opinion is represented by a numerical value, making it easier to organize and analyze the data.

Table 6.1: Difference between Measurement and Scaling:

Aspect	**Measurement**	**Scaling**
Definition	Assigning numbers or symbols to characteristics of objects based on specific rules.	Assigning numbers or meanings (semantics) to objects, typically statements about attitudes or opinions.
Purpose	Facilitates statistical analysis and communication of measurement rules and results.	Simplifies and organizes responses into numerical values for easier analysis of opinions or attitudes.
Focus	Measures characteristics of objects (e.g., feelings, opinions, attitudes).	Ranks or rates attitudes or opinions to represent them in a numeric format.
Example	Assigning ‘1’ for male and ‘2’ for female, or ‘1’ for “yes” and ‘2’ for “no” in a survey.	Assigning ‘1’ for “strongly agree,” ‘2’ for “somewhat agree,” and so on, to indicate levels of agreement.
Emphasis	Assigning numbers based on standardized and uniform rules.	Organizing responses to attitude statement for easier analysis and comparison.

6.3 LEVELS OF MEASUREMENT SCALES

The level of measurement refers to the relationship among the values that are assigned to the attributes, feelings or opinions for a variable. For example, the variable ‘whether the taste of fast food is good’ has a number of

attributes, namely, very good, good, neither good nor bad, bad and very bad. For the purpose of analysing the results of this variable, we may assign the values 1, 2, 3, 4 and 5 to the five attributes respectively. The level of measurement describes the relationship among these five values. Here, we are simply using the numbers as shorter placeholders for the lengthier text terms. We don't mean that higher values mean 'more' of something or lower values mean 'less' of something. We don't assume that 'good' which has a value of 2 is twice of 'very good' which has a value of 1. We don't even assume that 'very good' which is assigned the value '1' has more preference than 'good' which is assigned the value '2'. We simply use the values as a shorter name for the attributes, opinions, or feelings. The assigned values of attributes allow the researcher more scope for further processing of data and statistical analysis.

Typically, there are four levels of measurement scales or methods of assigning numbers: (a) Nominal scale, (b) Ordinal scale, (c) Interval scale, and (d) Ratio scale.

6.3.1 Nominal Scale

Nominal Scale is the crudest among all measurement scales but it is also the simplest scale. In this scale the different scores on a measurement simply indicate different categories. The nominal scale does not express any values or relationships between variables. For example, labeling men as '1' and women as '2' which is the most common way of labelling gender for data recording purposes does not mean that women are 'twice something or other' than men. Nor does it suggest that men are somehow 'better' than women. Another example of nominal scale is to classify the respondents on the basis of marital status "single", "married", "divorced", "widowed". The nominal scale is often referred to as a categorical scale. The assigned numbers have no arithmetic properties and act only as labels. The only statistical operation that can be performed on nominal scales is a frequency count. We can only determine mode. Other averages cannot be determined with the help of this scale.

In designing and developing a questionnaire, it is important that the response categories must include all possible responses. In order to have an exhaustive number of responses, you might have to include a category such as 'others', 'uncertain', 'don't know', or 'can't remember' so that the respondents will not distort their information by forcing their responses in one of the categories provided. You should also ensure that the categories provided are mutually exclusive so that they do not overlap or get duplicated in any way

6.3.2 Ordinal Scale

Ordinal Scale involves the ranking of items along the continuum of the characteristic being scaled/measured. In this scale, the items are classified according to whether they have more or less of a characteristic. For example, you may wish to ask the TV viewers to rank the TV channels according to their preference and the responses may look like this as given below in Table 6.2

Table 6.2: Example of ordinal scale

TV Channel	Viewers preferences
Channel 1	1
Channel 2	2
Channel 3	3
Channel 4	4

The main characteristic of the ordinal scale is that the categories have a logical or ordered relationship. This type of scale permits the measurement of degrees of difference, (that is, 'more' or 'less') but not the specific amount of differences (that is, how much 'more' or 'less'). This scale is very common in marketing, satisfaction and attitudinal research.

Another example is that a fast-food home delivery shop may wish to ask its customers:

How would you rate the service of our staff? (1) Excellent (2) Very Good (3) Good (4) Poor (5) Worst

Suppose respondent X gave the response 'Excellent' and respondent Y gave the response 'Good', we may say that respondent X thought that the service provided better than respondent Y to be thought. But we don't now how much better and even we can't say that both respondents have the same understanding of what constitutes 'good service'.

In marketing research, ordinal scales are used to measure relative attitudes, opinions, and preferences. Here we rank the attitudes, opinions and preferences from best to worst or from worst to best. However, the amount of difference between the ranks cannot be found out. Using ordinal scale data, we can perform statistical analysis like Median and Mode, but not the Mean.

6.3.3 Interval Scale

An interval Scale is a scale in which the numbers are used to rank attributes such that numerically equal distances on the scale represent equal distance in the characteristic being measured. An interval scale not only contains all the information of an ordinal scale, but it also allows us to compare the difference/distance between attributes. For example, the difference between '1' and '2' is equal to the difference between '3' and '4'. Further, the difference between '2' and '4' is twice the difference between '1' and '2'. However, in an interval scale, the zero point is arbitrary and does not represent absolute or true zero. This, of course, has implications for the type of data manipulation and analysis that can be performed on data collected in this form. It is possible to add or subtract a constant to all of the scale values without affecting the form of the scale but one cannot multiply or divide the values. Measuring temperature is an example of interval scale. We cannot say 40^0C is twice as hot as 20^0C. The reason for this is that 0^0C does not mean that there is no temperature, but a relative point on the Centigrade Scale. Due

to the lack of an absolute zero point, the interval scale does not allow the conclusion that 40°C is twice as hot as 20°C.

Interval scales may be either in numeric or semantic formats. The following are two more examples of interval scales one in numeric format and another in semantic format in Table 6.3 and Table 6.4 respectively.

i) **Example of Interval Scale in Numeric Format**

Table: 6.3 Interval Scale in Numeric Format

Food supplied is:	1	2	3	4	5	*Indicate your score on the concerned blank and circle the appropriate number on each line.*
Fresh					✓	
Tastes good					✓	
Value for money			✓			
Attractive packaging				✓		
Prompt time delivery		✓				

ii) **Example of Interval Scale in Semantic Format**

Please indicate your views on the food supplied by ABC Fast Food Shop by scoring them on a five points scale from 1 to 5 (that is, 1=Excellent, 2=Very Good, 3=Good, 4=Poor, 5=Worst). Indicate your views by ticking the appropriate responses below:

Table 6.4: Interval Scale in Semantic Format

Food supplied is:	**Excellent**	**Very Good**	**Good**	**Poor**	**Worst**
Fresh					
Tastes good					
Value for money					
Attractive packaging					
Prompt time delivery					

The interval scales allow the calculation of averages (Mean, Median, Mode) and dispersion measures like Range and Standard Deviation.

6.3.4 Ratio Scale

Ratio Scale is the highest level of measurement scales. This has the properties of an interval scale together with a fixed (absolute) zero point. The absolute zero point allows us to construct a meaningful ratio. Examples of ratio scales include weights, lengths and times. In the marketing research, most counts are ratio scales. For example, the number of customers of a bank's ATM in the last three months is a ratio scale. This is because you can compare this with the previous three months. Ratio scales permit the

researcher to compare both differences in scores and relative magnitude of scores. For example, the difference between 10 and 15 minutes is the same as the difference between 25 and 30 minutes and 30 minutes is twice as long as 15 minutes. Most financial research that deals with rupee values utilizes ratio scales. However, for most behavioural research, interval scales are typically the highest form of measurement. Most statistical data analysis procedures do not distinguish between the interval and ratio properties of the measurement scales and it is sufficient to say that all the statistical operations that can be performed on interval scale can also be performed on ratio scales.

Now you must be wondering why you should know the level of measurement. Knowing the level of measurement helps you to decide on how to interpret the data. For example, when you know that a measure is nominal then you know that the numerical values are just short codes for longer textual names. Also, knowing the level of measurement helps you to decide what statistical analysis is appropriate on the values that were assigned. For example, if you know that a measure is nominal, then you would not need to find the mean of the data values or perform a t-test on the data.

It is important to recognise that there is a hierarchy implied in the levels of measurement. At lower levels of measurement, assumptions tend to be less restrictive and data analyses tend to be less sensitive. At each level up the hierarchy, the current level includes all the qualities of the one below it and adds something new. In general, it is desirable to have a higher level of measurement (that is, interval or ratio) rather than a lower one (that is, nominal or ordinal). Figure 6.1 shows examples of different types of Scale.

Table 6.5: Types of scales and their Characteristics

Scale	Basic Characteristics	Common Examples	Marketing Examples	Descriptive Statistics	Inferential Statistics
Nominal	Numbers identify & classify objects	Social Security no., numbering of football players	Brand no. store types	Percentage, mode	Chi-square binomial test
Ordinal	No. indicate the relative positions of objects but not the magnitude of differences between them	Quality rankings, rankings of teams in a tournament	Preference rankings, market position, social class	Percentile, median	Rank-order correlation, Friedman ANOVA
Interval	Difference between objects	Temperature (Fahrenheit)	Attitudes, opinions, index	Range, mean, standard Deviation	Product-moment
Ratio	Zero point is fixed, ratios of scale values can be compared	Length, weight	Age, sales, income, costs	Geometric mean, harmonic mean	Coefficient of variation

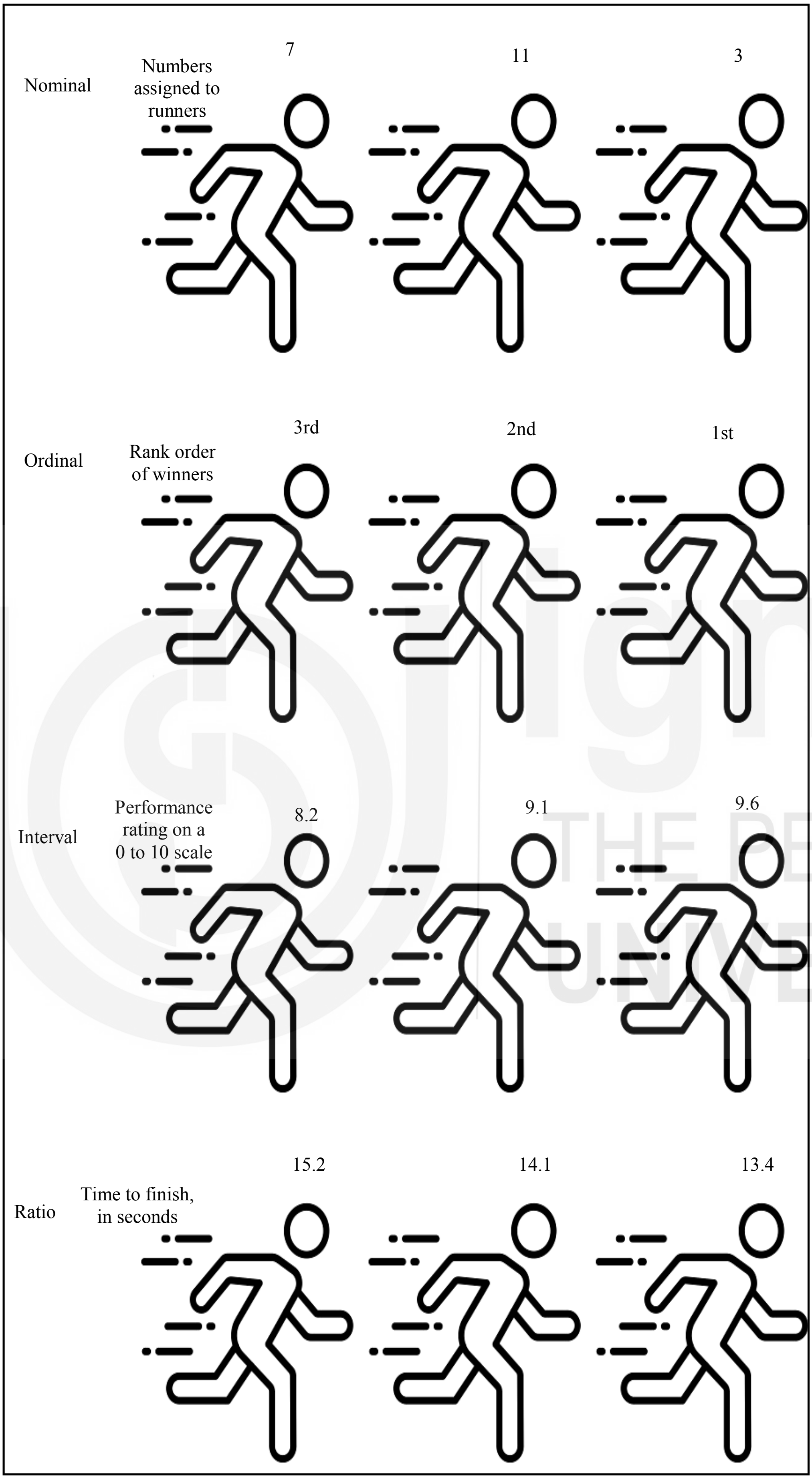


Figure 6.1: Example of Nominal, Ordinal, Scale, Ratio

Check Your Progress-A

1) Compare and contrast nominal, ordinal, interval, and ratio scales with examples.

....

....

....

....

....

2) Why should the researcher know the level of measurement?

....

....

....

....

....

3) What are the main statistical limitations of nominal scale?

....

....

....

....

....

4) Mark the most correct answer

   i) What does measurement focus on?

      a) Measuring respondents directly

      b) Assigning numbers to objects' characteristics

      c) Analyzing survey results manually

      d) Creating a new set of scaling rules

   ii) Which scale includes an absolute zero point?

      a) Nominal

      b) Ordinal

      c) Interval

      d) Ratio

   iii) What is a primary characteristic of the interval scale?

      a) It uses absolute zero points.

      b) It ranks items but doesn't quantify differences.

      c) It maintains equal distances between scale points.

      d) It assigns labels without implying order

6.4 TYPES OF SCALES TECHNIQUES

The various types of scaling techniques used in research can be classified into two categories: (a) Comparative Scales, and (b) Non-Comparative Scales. In comparative scaling, the respondent is asked to compare one object with another. For example, the researcher can ask the respondents whether they prefer brand A or brand B of a detergent. On the other hand, in non-comparative scaling respondents need only evaluate a single object. Their evaluation is independent of the other object which the researcher is studying. Respondents using a non-comparative scale employ whatever rating standard seems appropriate to them. The comparative scales can further be divided into the following four types of scaling techniques: (a) Paired Comparison Scale, (b) Rank Order Scale, (c) Constant Sum Scale, and (d) Q-sort Scale. Non-comparative techniques consist of continuous and itemized rating scales. Figure 6.2 shows the classification of these scaling techniques.

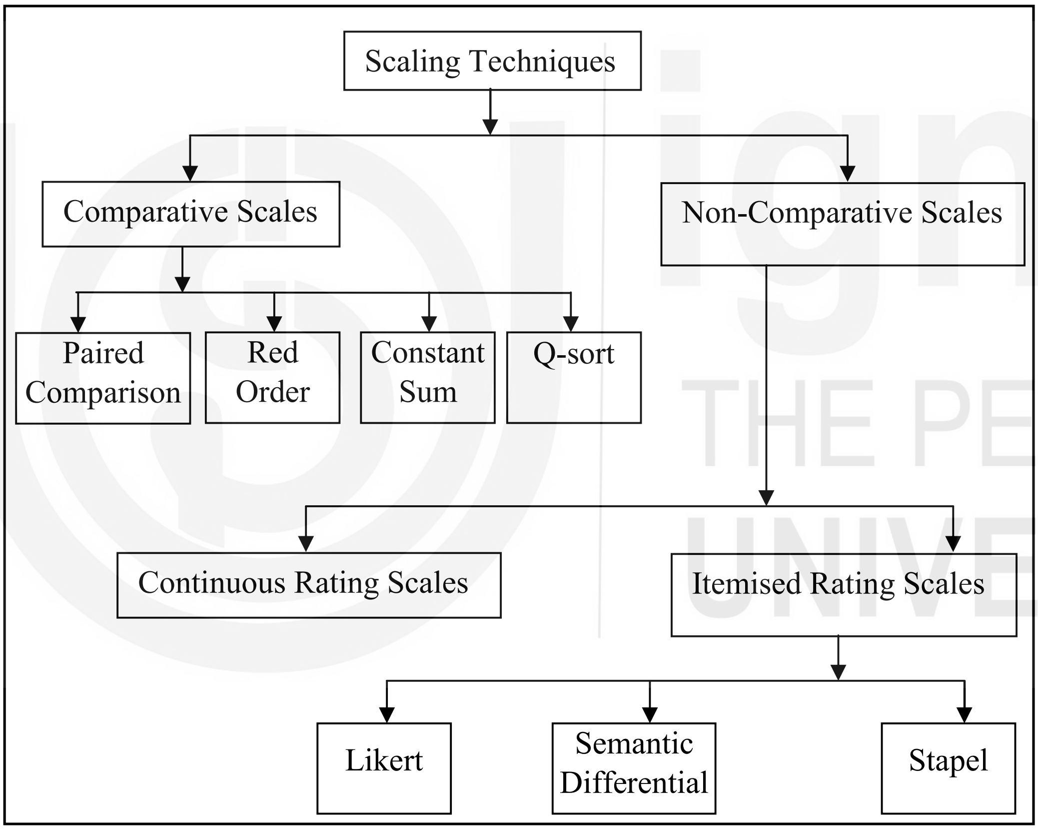


Figure 6.2: Scaling Techniques

6.4.1 Comparative Scales

Comparative scales involve the direct comparison of stimulus objects. For example, respondents may be asked whether they prefer Coke or Pepsi. The comparative scales can further be divided into the following four types of scaling techniques: (a) Paired Comparison Scale, (b) Rank Order Scale, (c) Constant Sum Scale, and (d) Q-sort Scale.

6.4.1.1 Paired Comparison

Paired Comparison Scale: This is a comparative scaling technique in which a respondent is presented with two objects at a time and asked to select one object (rate between two objects at a time) according to some criterion. The data obtained are ordinal in nature. For example, there are four types of cold drinks - Coke, Pepsi, Sprite, and Limca. The respondents can prefer Pepsi to Coke or Coke to Sprite, etc. In all we can have the following six comparisons.

Coke–Pepsi

Coke–Sprite

Coke–Limca

Pepsi–Sprite

Pepsi–Limca

Sprite–Limca

In general, with n brands we have N(N-1)/2 paired comparisons. The following is the data recording format using the paired comparisons.

Table 6.6: Paired Comparison Scale of cold drink

Brand	Coke	Pepsi	Sprite	Limca
Coke	—	√		
Pepsi		—		
Sprite	√	√	—	
Limca	√	√	√	—
No. of times preferred	2	3	1	0

A √ in a particular box means that the brand in that column was preferred over the brand in the corresponding row. In the above recording, Coke was preferred over Sprite, Coke over Limca, in this case the number of times coke preferred was 2 times. Similarly, Pepsi over Coke, Pepsi over Sprite, Pepsi over Limca, in this case Pepsi was 3 time preferred. Thus, the number of times a brand was preferred is obtained by summing the √ s in each column.

The following table gives paired comparison of data (assumed) for four brands of cold drinks.

Table 6.7: Paired Comparison of cold drink

Brand	Coke	Pepsi	Sprite	Limca
Coke	–	0.90	0.64	0.14
Pepsi	0.10	–	0.32	0.02
Sprite	0.36	0.68	–	0.15
Limca	0.86	0.98	0.85	–

The entries in the boxes represent the proportion of respondents preferring 'column brand' and to 'row' brand. For example, 90% prefer Pepsi to Coke and only 10% prefer Coke to Pepsi, etc.

Paired comparison is useful when the number of brands is limited, since it requires direct comparison and overt choice. One of the disadvantages of paired comparison scale is violation of the assumption of transitivity may occur. For example, in our example (Table 6.6) the respondent preferred Coke 2 times, Pepsi 3 times, Sprite 1 time, and Limca 0 times. That means, preference-wise, Pepsi >Coke, Coke >Sprite, and Sprite >Limca. However, the number of times Sprite was preferred should not be that of Coke. In other words, if A>B and B >C then C >A should not be possible. Also, the order in which the objects are presented may bias the results. The number of items/brands for comparison should not be too many. As the number of items increases, the number of comparisons increases geometrically. If the number of comparisons is too large, the respondents may become fatigued and no longer be able to carefully discriminate among them. The other limitation of paired comparison is that this scale has little resemblance to the market situation, which involves selection from multiple alternatives. Also, respondents may prefer one item over certain others, but they may not like it in an absolute sense.

6.4.1.2 Rank Order

Rank Order Scale: This is another type of comparative scaling technique in which respondents are presented with several items simultaneously and asked to rank them in the order of priority. This is an ordinal scale that describes the favoured and unfavoured objects, but does not reveal the distance between the objects. For example, if you are interested in ranking the preference of some selected brands of cold drinks, you may use the following format for recording the responses.

Instructions: Rank the following brands of cold drinks in order of preference. Begin by picking out the one brand you like most and assign it a number1. Then find the second most preferred brand and assign it a number 2. Continue this procedure until you have ranked all the brands of cold drinks in order of preference. The least preferred brand should be assigned a rank of 4. Also remember no two brands receive the same rank order.

Format:

Table 6.8: Preference of cold drink brands using rank order scaling

Brand	Rank
(a) Coke	3
(b) Pepsi	1
(c) Limca	2
(d) Sprite	4

Like paired comparison, the rank order scale, is also comparative in nature. The resultant data in rank order is ordinal data. This method is more realistic

in obtaining the responses and it yields better results when direct comparison is required between the given objects. The major disadvantage of this technique is that only ordinal data can be generated.

6.4.1.3 Constant Sum

Constant Sum Scale: In this scale, the respondents are asked to allocate a constant sum of units such as points, rupees, or chips among a set of stimulus objects with respect to some criterion. For example, you may wish to determine how important the attributes of price, fragrance, packaging, cleaning power, and lather of a detergent are to consumers. Respondents might be asked to divide a constant sum to indicate the relative importance of the attributes using the following format.

Instructions: Between attributes of detergent please allocate 100 points among the attributes so that your allocation reflects the relative importance you attach to each attribute. The more points an attribute receives, the more important the attribute is. If an attribute is not at all important, assign it zero points. If an attribute is twice as important as some other attribute, it should receive twice as many points.

Format:

Table 6.9: Importance of detergent attributes using a constant sum scale

Attribute	Number of Points
(a) Price	50
(b) Fragrance	05
(c) Packaging	10
(d) Cleaning Power	30
(e) Lather	05
Total Points	100

"If an attribute is assigned a higher number of points, it would indicate that the attribute is more important." From the above Table, the price of the detergent is the most important attribute for the consumers followed by cleaning power, packaging. Fragrance and lather are the two attributes that the consumers cared about the least but preferred equally." The advantage of this technique is saving time. However, there are two main disadvantages. The respondents may allocate more or fewer points than those specified. The second problem is rounding off error if too few attributes are used and the use of a large number of attributes may be too taxing on the respondent and cause confusion and fatigue.

Q-sort Scaling

This is a comparative scale that uses a rank order procedure to sort objects based on similarity with respect to some criterion. The important characteristic of this methodology is that it is more important to make

comparisons among different responses of a respondent than the responses between different respondents. Therefore, it is a comparative method of scaling rather than an absolute rating scale. In this method the respondent is given statements in a large number for describing the characteristics of a product or a large number of brands of a product. For example, you may wish to determine the preference from among a large number of magazines. The following format shown in Table 6.9 may be given to a respondent to obtain the preferences.

Instructions: The bag given to you contain pictures of 90 magazines. Please choose 10 magazines you 'prefer most', 20 magazines you 'like', 30 magazines to which you are 'neutral (neither like nor dislike)', 20 magazines you 'dislike', and 10 magazines you 'prefer least'. Please list the sorted magazine names in the respective columns of the form provided to you.

Format:

Table 6.9: Preference of Magazines Using Q-Sort Scale Procedure

Prefer Most	Like	Neutral	Dislike	Prefer Least
……………..	……………..	……………..	……………..	……………..
……………..	……………..	……………..	……………..	……………..
……………..	……………..	……………..	……………..	……………..
……………..	……………..	……………..	……………..	……………..
……………..	……………..	……………..	……………..	……………..
……………..	……………..	……………..	……………..	……………..
……………..	……………..	……………..	……………..	……………..
……………..	……………..	……………..	……………..	……………..
……………..	……………..	……………..	……………..	……………..
……………..	……………..	……………..	……………..	……………..
……………..	……………..	……………..	……………..	……………..
(10)	……………..	……………..	……………..	(10)
	……………..	……………..	……………..	
	……………..	……………..	……………..	
	……………..	……………..	……………..	
	……………..	……………..	……………..	
	……………..	……………..	……………..	
	……………..	……………..	……………..	
	……………..	……………..	……………..	
	……………..	……………..	……………..	
	……………..	……………..	……………..	
	(20)	……………..	(20)	
		……………..		
		……………..		
		……………..		
		……………..		
		……………..		
		……………..		
		……………..		
		……………..		
		……………..		
		(30)		

Note that the number of responses to be sorted should not be less than 60 or not more than 140. A reasonable range is 60 to 90 responses that result in a normal or quasi-normal distribution. This method is faster and less tedious than paired comparison measures. It also forces the subject to conform to quotas at each point of scale so as to yield a quasi-normal distribution. The utility of Q-sort in marketing research is to derive clusters of individuals who display similar preferences, thus representing unique market segments.

6.4.2 Non-Comparative Scales

Respondents using a non-comparative scale employ whatever rating standard seems appropriate to them. They do not compare the object being rated either with another object or to some specified standard, such as 'your ideal brand'. They evaluate only one object at a time. Non-comparative scaling techniques can be further divided into: (a) Continuous Rating Scale, and (b) Itemised Rating Scale.

6.4.1.1 Continuous Rating scales

a) **Continuous Rating Scales:** It is very simple and highly useful. In the continuous rating scale, the respondents' rates the objects by placing a mark at the appropriate position on a continuous line that runs from one extreme of the criterion variable to the other. Examples of continuous rating scale are given below Figure 6.3.

Question: How would you rate the TV advertisement as a guide for buying?

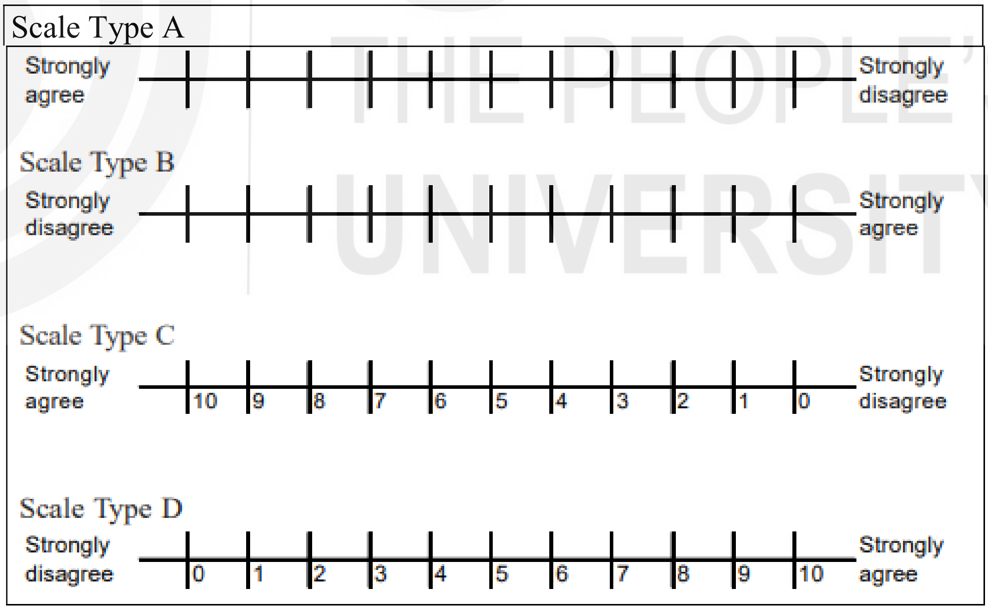


Figure 6.3 Continuous Rating Scale

When scale type A and B are used, the respondents score is determined either by dividing the line into as many categories as desired and assigning the respondent a score based on the category into which his/her mark falls, or by measuring distance, in millimeters, centimeters, or inches from either end of the scale. Whichever of the above continuous scale is used, the results are normally analyzed as interval scaled.

6.4.2.2 Itemized rating Scale

Itemised rating scale is a scale having numbers or brief descriptions associated with each category. The categories are ordered in terms of scale position and the respondents are required to select one of the limited number of categories that best describes the product, brand, company, or product attribute being rated. Itemised rating scales are widely used in marketing research.

The itemised rating scales can be in the form of: (a) graphic, (b) verbal, or (c) numeric as shown in below Figure 6.4.

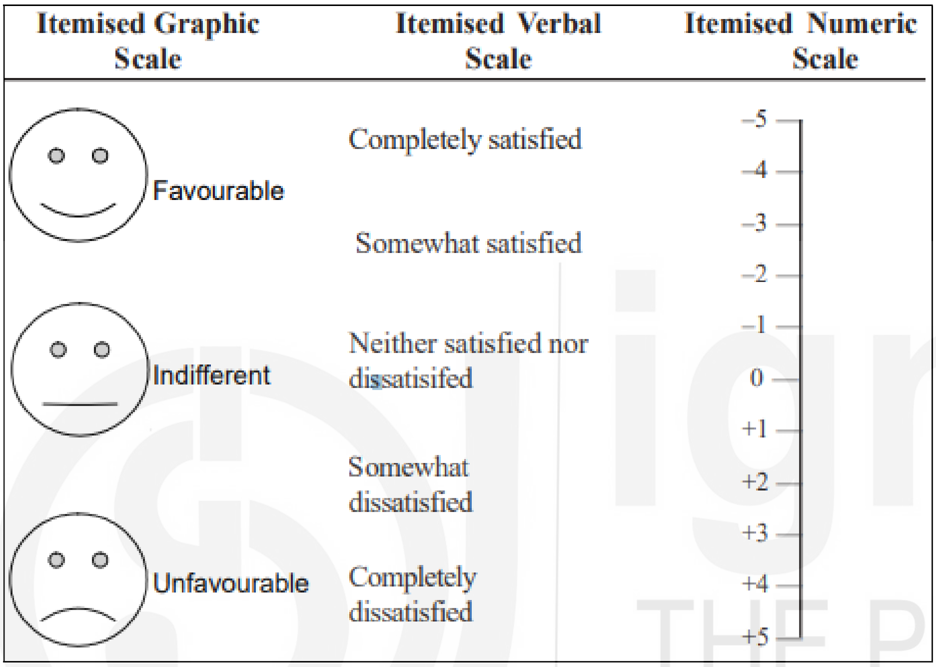


Figure: 6.4: Graphic, Verbal and Numeric Itemised Scale

Some rating scales may have only two response categories such as: agree and disagree. Inclusion of more response categories provides the respondent more flexibility in the rating task. Consider the following questions:

1. How often do you visit the supermarket located in your area of residence?

   • Never, • Rarely, • Sometimes, • Often, • Very often

2. In your case, how important is the price of brand X shoes when you buy them?

   • Very important, • Fairly important, • Neutral, • Not so important

Each of the above category scales is a more sensitive measure than a scale with only two responses since they provide more information.

Wording is an extremely important factor in the usefulness of itemised scales. Table 6.11 shows some common wordings for categories used in itemised scales.

Table 6.11: Some common words for categories used in Itemised Rating scales

Quality:				
Excellent	Good	Not decided	Poor	Worst
Very Good	Good	Neither good not bad	Fair	Poor
Importance:				
Very Important	Fairly important	Neutral	Not so Important	Not at all important
Interest:				
Very interested	Somewhat interested	Neither interested nor disinterested	Somewhat uninterested	Not very interested
Satisfaction:				
Completely satisfied	Somewhat satisfied	Neither satisfied nor dissatisfied	Somewhat dissatisfied	Completely dissatisfied
Frequency:				
All of the time	Very often	Often	Sometimes	hardly ever
Very often	often	Sometimes	Rarely	Never
Truth:				
Very true	Somewhat true	Not very true	Not at all true	
Purchase Interest:				
Definitely will buy	Probably will buy	Probably will not buy	Definitely will not buy	
Level of Agreement:				
Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree
Dependability:				
Completely dependable	Somewhat dependable	Not very dependable	Not at all dependable	
Style:				
Very stylish	Somewhat stylish	Not very stylish	Completely unstylish	
Cost:				
Extremely expensive	Expensive	Neither expensive nor inexpensive	Slightly inexpensive	Very inexpensive
Ease of use:				
Very case to use	Somewhat easy to use	Not very easy to use	Difficult to use	
Modernity:				
Very modern	Somewhat modern	Neither modern nor old-fashioned	Somewhat old fashioned	Very old fashioned
Alert:				
Very alert	Alert	Not alert	Not at all alert	

In this section we will discuss most commonly used itemised rating scales, namely (a) Likert scale, (b) Semantic Differential Scale, and (c) Stapel Scale.

6.4.2.2.1 Likert

In business research, the Likert scale, developed by Rensis Likert, is extremely popular for measuring attitudes, because the method is simple to administer. With the Likert scale, the respondents indicate their own attitudes by checking how strongly they agree or disagree with carefully worded statements that range from very positive to very negative towards the attitudinal object. Respondents generally choose from five alternatives (say strongly agree, agree, neither agree nor disagree (not sure), disagree, strongly disagree). Consider the following example of a study or measuring attitudes towards cricket.

Table 6.12: Likert Scale for measuring attitude towards cricket

	Strongly Agree	Agree	Not Sure	Disagree	Strongly disagree
It is more fun to play a tough, competitive cricket match than to play an easy one.	5	4	3	2	1

To measure the attitude, the researchers assign weights or scores to the alternative responses. In the above example the scores 5 to 1 are assigned to the responses. Strong agreement of the respondent indicates the most favorable attitudes on the statement, and the score 5 is assigned to it. On the other hand, strong disagreement of the respondent indicates the most unfavorable attitude on the statement, and the score 1 is assigned to it. If a negative statement towards the object is given, the corresponding scores would be reversed. In this case, the response 'strongly agree' will get a score of 1 and the response 'strongly disagree' will get a score of 5.

A Likert scale may include a number of items or statements. Each statement is assumed to represent an aspect of an attitudinal domain. For example, Table 6.13 shows the items in a Likert Scale to measure opinions on food products.

Table 6.13: A Likert Scale for studying opinions on food products.

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
If the price of raw materials fall, firms too should reduce 5 the price of the food products.	1	2	3	4	5
There should be uniform price throughout the country for food products	1	2	3	4	5
The food companies should concentrate more on keeping hygiene while manufacturing food products.	1	2	3	4	5
The expiry dates should be printed on the food products before they are delivered to consumers in the market.	1	2	3	4	5
There should be government regulations on the firms in keeping acceptable quality and on the prices	1	2	3	4	5
Now-a-days most food companies are concerned only with profit making rather than taking care of quality.	1	2	3	4	5

Each respondent is asked to circle his opinion on a score against each statement. The final score for the respondent on the scale is the sum of their ratings for all the items. The very purpose of Likert's Scale is to ensure the final items evoke a wide response and discriminate among those with positive and negative attitudes. Items that are poor (because they lack clarity or elicit mixed response patterns) are detected from the final statement list. This will ensure us to discriminate between high positive scores and high negative scores. However, many business researchers do not follow this procedure and you may not be in a position to distinguish between high positive scores and high negative scores because all scores look alike. Hence a disadvantage of the Likert Scale is that it is difficult to know what a single summated score means. Many patterns of response to the various statements can produce the same total score. The other disadvantage of Likert Scale is that it takes longer time to complete than other itemised rating scales because respondents have to read each statement. Despite the above disadvantages, this scale has several advantages. It is easy to construct, administer and use.

6.4.2.2.2 Semantic Differential Scale

This is a seven-point rating scale with endpoints associated with bipolar labels (such as good and bad, complex and simple) that have semantic meaning. The Semantic Differential scale is used for a variety of purposes. It can be used to find whether a respondent has a positive or negative attitude towards an object. It has been widely used in comparing brands, products and company images. It has also been used to develop advertising and promotion strategies and in a new product development study. Look at the below Table 6.14, and 6.15 and Figure 6.5 for examples of Semantic Differential Scale.

Table 6.14: Examples of Semantic Differential Scale

Modern	—	—	—	—	—	—	—	Old-fashioned
Good	—	—	—	—	—	—	—	Bad
Clean	—	—	—	—	—	—	—	Dirty
Important	—	—	—	—	—	—	—	Unimportant
Expensive	—	—	—	—	—	—	—	Inexpensive
Useful	—	—	—	—	—	—	—	Useless
Strong	—	—	—	—	—	—	—	Weak
Quick	—	—	—	—	—	—	—	Slow

In the Semantic Differential Scale only extremes have names. The extreme points represent the bipolar adjectives with the central category representing the neutral position. The in between categories have blank spaces. A weight is assigned to each position on the scale. The weights can be such as +3, +2, +1, 0, –1, –2, –3 or 7,6,5,4,3,2,1. The following is an example of Semantic Differential Scale to study the experience of using a particular brand of body lotion see Table 6.15

Table 6.15: Semantic differential effect for Body Lotion

In my experience, the use of body lotion of Brand-X was:

	+3	+2	+1	0	-1	-2	-3	
Useful	—	—	—	—	—	—	—	Useless
Attractive	—	—	—	—	—	—	—	Unattractive
Passive	—	—	—	—	—	—	—	Active
Beneficial	—	—	—	—	—	—	—	Harmful
Interesting	—	—	—	—	—	—	—	Boring
Dul	—	—	—	—	—	—	—	Sharp
Pleasant	—	—	—	—	—	—	—	Unpleasant
Cold	—	—	—	—	—	—	—	Hot
Good	—	—	—	—	—	—	—	Bad
Likable	—	—	—	—	—	—	—	Unlikable

In the Semantic Differential Scale, the phrases used to describe the object form a basis for attitude formation in the form of positive and negative phrases. The negative phrase is sometimes put on the left side of the scale and sometimes on the right side. This is done to prevent a respondent with a positive attitude from simply checking the left side and a respondent with a negative attitude checking on the right side without reading the description of the words.

The respondents are asked to check the individual cells depending on the attitude. Then one could arrive at the average scores for comparisons of different objects. The following Figure 6.5 shows the experiences of 100 consumers on 3 brands of body lotion.

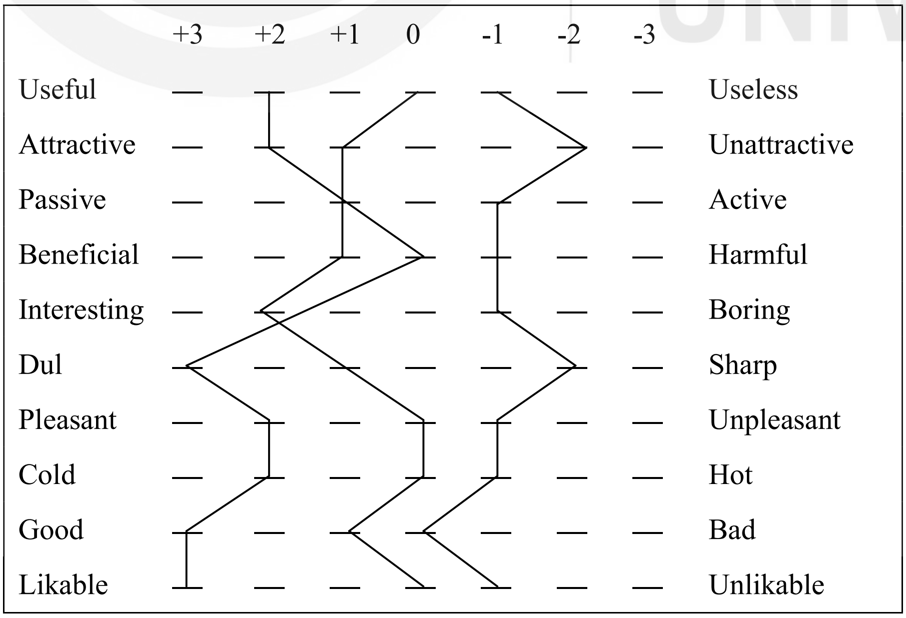


Figure:6.5: Semantic differential scale for 3 body lotion brands

In the above example, first the individual respondent scores for each dimension are obtained and then the average scores of all 100 respondents, for each dimension and for each brand were plotted graphically. The maximum score possible for each brand is + 30 and the minimum score possible for each brand is –30. Brand-X has score +14. Brand-Y has score +7, and Brand-Z has score –11. From the scale we can identify which phrase needs improvement for each Brand. For example, Brand-X needs to be improved upon benefits and Brand-Y on pleasantness, coldness and likeability. Brand Z needs to be improved on all the attributes.

6.4.2.2.3 Stapel Scale

The Stapel scale was originally developed to measure the direction and intensity of an attitude simultaneously. Modern versions of the stapel scale place a single adjective as a substitute for the semantic differential when it is difficult to create pairs of bipolar adjectives. The modified stapel scale places a single adjective in the centre of an even number of numerical values (say, +3, +2, +1, 0, –1, –2, –3). This scale measures how close to or how distant from the adjective a given stimulus is perceived to be. The following is an example of a staple scale.

Instructions: Select a positive number for words that you think describe personal banking of a bank accurately. The more accurately you think the word describes the bank, the larger the plus number you should choose. Select a negative number for words you think do not describe the bank accurately. The less accurately you think the word describes the bank, the larger the minus number you should choose.

Format:

Table 6.16: Staple scale

+5	+5
+4	+4
+3	+3
+2	+2
+1	+1
Friendly Personnel	Competitive Loan Rates
–1	–1
–2	–2
–3	–3
–4	–4
–5	–5

The following format shows an example of stapel Scale that illustrates respondents' description on personal banking of a bank.

	+4	+3	+2	+1	-1	-2	-3	-4
Fast Service	—	—	—	—	—	—	—	—
Friendly	—	—	—	—	—	—	—	—
Honest	—	—	—	—	—	—	—	—
Convenient Location	—	—	—	—	—	—	—	—
Convenient Hours	—	—	—	—	—	—	—	—
Dul	—	—	—	—	—	—	—	—
Good Services	—	—	—	—	—	—	—	—
High Saving Rates	—	—	—	—	—	—	—	—

Figure 6.6: Opinion of respondent on one Bank Services

Each respondent is asked to circle his opinion on a score against each phrase that describes the object. The final score of the respondent on a scale is the sum of their ratings for all the items. Also, the average score for each phrase is obtained by totaling the final score of all the respondents for that phrase divided by the number of respondents of the phrase. The following Figure shows the opinions of 100 respondents on two banks.

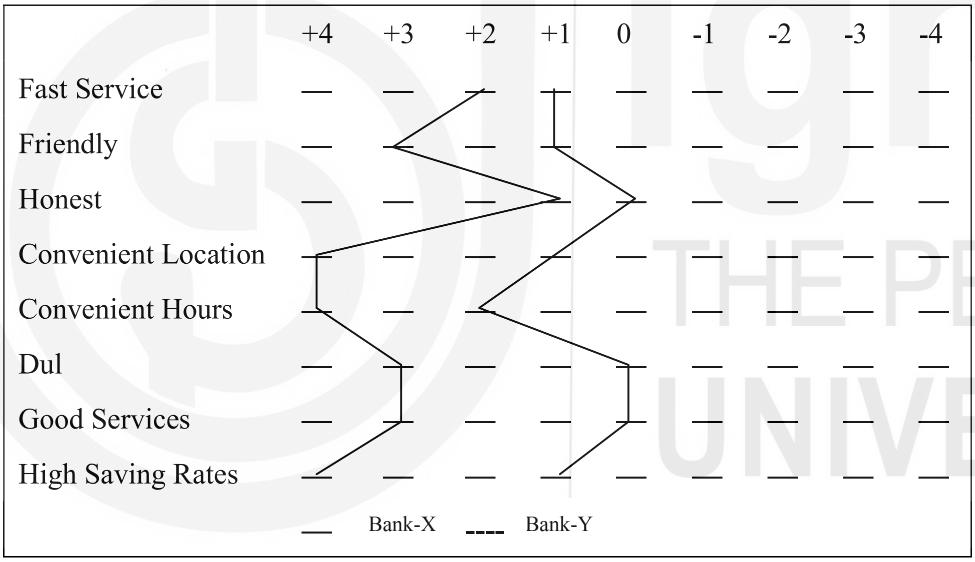


Figure 6.7: Opinion of respondent on two Bank Services

In the above example first the individual respondent's scores for each phrase that describes the selected bank are obtained and then the average scores of all 100 respondents for each phrase are plotted graphically. The maximum score possible for each bank is +32 and the minimum possible score for each brand is –32. In the example, Bank-X has score +24, and Bank-Y has score +3. From the scale we can identify which phrase needs improvement for each Bank. The advantages and disadvantages of the stapel scale are very similar to those for the semantic differential scale. However, the stapel scale tends to be easier to construct and administer, especially over telephone, since the stapel scale does not call for the bipolar adjectives as does the semantic differential scale. However, research on comparing the stapel scale with semantic differential scale suggests that the results of both the scales are largely the same.

6.5 RELIABILITY AND VALIDITY OF A SCALE

In the previous section we have learned that scales are important tools for translating abstract and often intangible concepts, such as attitudes, perceptions, or behaviors, into quantifiable and interpretable data. It is because these abstract concepts are not measured directly. These abstract constructs, which are inherently complex and subjective, require careful measurement to ensure that the data collected can be analyzed meaningfully. For instance, measuring something as abstract as "customer satisfaction" or "emotional intelligence" demands a well-designed scale that captures the essence of the construct in a measurable form.

The effectiveness of a scale depends on its ability to consistently and accurately measure the concept it is intended to represent. This involves not only capturing the construct in its entirety but also doing so in a way that minimizes errors and biases. This is where reliability and validity become critical. These two principles form the backbone of scale evaluation, acting as benchmarks for determining its quality. They ensure that the scale provides results that are both consistent across various conditions (reliability) and accurately reflective of the construct being measured (validity). Refer Figure 6.8.

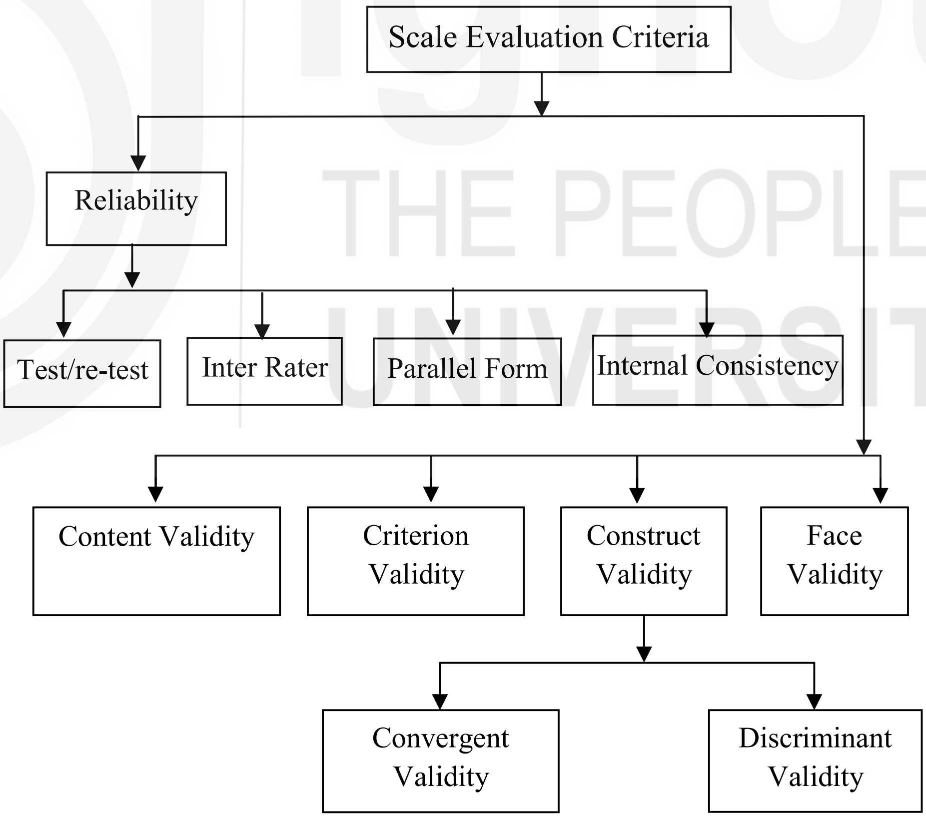


Figure 6.8: Scale Evaluation Criteria

A scale that is both reliable and valid allows researchers to draw conclusions with confidence. Reliability ensures that the measurement is consistent over time, across different raters, or between different versions of the scale. On the other hand, validity ensures that the scale measures what it is supposed to measure. We will discuss details on reliability and validity in the below section.

6.5.1 Reliability of Scale

Reliability refers to the extent to which a scale produces consistent results if repeated measurements are made. A reliable instrument produces the same results under consistent conditions, minimizing random errors.

Types of Reliability in a Scale:

a) **Test-Retest Reliability**: Test-retest reliability evaluates the stability of a scale over time by measuring whether it produces consistent results when administered to the same group of respondents under the same conditions at different points in time. Purpose of this Test-Retest ability is to check if the scale can reliably measure the construct across time intervals.

   For example: A customer satisfaction survey given to a group of participants twice, with a two-week gap, should yield similar results if the construct (customer satisfaction) remains unchanged.

b) **Inter-Rater Reliability**: Inter-rater reliability assesses the degree of agreement or consistency between different observers or raters using the same scale. The purpose of inter-rater reliability is to assesses the degree of agreement or consistency between different observers or raters using the same scale.

   For example: Two interviewers independently scoring a candidate's performance on a behavioral assessment scale should arrive at similar scores if the scale is reliable.

c) **Parallel-Forms Reliability:** Parallel-form reliability compares the consistency of results between two equivalent but different versions of a scale designed to measure the same construct. The purpose of parallel-form-reliability is to verify that the scale produces reliable results irrespective of the version used.

   For example: Two versions of a language proficiency test, with equivalent but different questions, should yield similar scores for the same group of test-takers.

d) **Internal Consistency Reliability:** Internal consistency reliability examines the degree to which items within a scale correlate with each other and measure the same construct. Purpose of internal consistency reliability is to ensure that all items in a test contribute to measuring a unified concept.

For example: In a job satisfaction survey, questions about work environment, management, and peer relationships should all align and contribute to the overall concept of job satisfaction.

Assessment Methods:

Cronbach's Alpha: A common measure of internal consistency; values above 0.7 are generally considered acceptable.

Split-Half Reliability: Divides the scale into two halves (e.g., odd and even items) and checks the consistency of results between them.

6.5.2 Validity of Scale

Validity refers to the degree to which a scale accurately measures the concept or construct it is intended to measure. It is the extent to which a measurement represents characteristics that exist in the phenomenon under investigation. Validity ensures that the data collected is meaningful, relevant, and reflective of the underlying theoretical framework. A scale lacking validity, even if reliable, can produce results that are systematically incorrect, leading to flawed interpretations and conclusions. Therefore, establishing the validity of a scale is a critical step in research design and scale development.

Types of Validity in Scales:

a) **Content Validity:** Content validity evaluates whether the scale comprehensively covers all dimensions and aspects of the construct it intends to measure. It ensures that the scale captures the full scope of the concept without omitting key elements. The purpose of content validity is to determine whether the items in the scale are representative of the construct.

   For Example**:** A scale measuring leadership qualities should include items that assess decision-making, communication skills, vision-setting, conflict resolution, and team-building. Missing any of these aspects could lead to an incomplete assessment.

   To establish content validity the researcher should consult subject matter experts and do proper review of literature to ensure that all components or dimensions of construct have been considered.

b) **Construct Validity:** Construct validity ensures that the scale measures the theoretical construct it is intended to measure. It examines whether the scale aligns with established theoretical expectations and related constructs. The purpose of construct validity is to evaluate the degree to which the scale reflects the theoretical framework and discriminates between related but distinct constructs.

   For example: A scale designed to measure emotional intelligence should demonstrate strong correlations with constructs like empathy, interpersonal skills, and emotional regulation while showing low correlations with unrelated constructs like mechanical aptitude.

   To establish Construct validity researcher should ensure convergent and discriminant validity

   Convergent validity: It ensures that the scale correlates positively with measures of similar constructs (e.g., emotional intelligence correlating with interpersonal skills).

   Discriminant validity: It confirms that the scale does not correlate strongly with measures of unrelated constructs (e.g., emotional intelligence should not correlate with numerical ability).

c) **Criterion Validity:** It determines how well the scale's results relate to an external benchmark or standard, assessing its ability to predict or

correlate with outcomes. The purpose is to validate the scale by comparing it with established criteria or outcomes. It includes:

Predictive validity: It Assesses how well the scale predicts future outcomes. For example: A job aptitude test should predict the candidate's future job performance.

Concurrent Validity: Measures the agreement between the scale and an existing validated instrument at the same time. For Example: A newly developed mental health scale should show strong correlations with an established mental health questionnaire.

d) **Face Validity:** Face validity is the simplest and most basic type of validity. It refers to whether a scale looks like it measures what it is supposed to measure, just by a quick glance or based on first impressions

6.6 SELECTION OF AN APPROPRIATE SCALING TECHNIQUE

In this unit, so far, you have learnt some of the important scaling techniques that are frequently used in attitudinal research for the measurement of attitudes. Each of these techniques has some advantages and disadvantages. Now you may ask which technique is more appropriate to use to measure attitudes. Virtually any technique can be used to measure attitudes. But at the same time all techniques are not suitable for all purposes. As a general rule, you should use a scaling technique that will yield the highest level of information feasible in a given situation. Also, if possible, the technique should permit you the use of a variety of statistical analysis. A number of issues decide the choice of scaling technique. Some significant issues are:

1) **Problem Definition and Statistical Analysis:** The Choice between ranking, sorting, or rating techniques is determined by the problem definition and the type of statistical analysis likely to be performed. For example, ranking provides only ordinal data that limits the use of statistical techniques.

2) **The Choice between Comparative and Non-comparative Scales:** Sometimes it is better to use a comparative scale rather than a non-comparative scale. Consider *the following example:*

How satisfied are you with the brand- X detergent that you are presently using?

Completely satisfied	*Somewhat satisfied*	*Neither satisfied nor dissatisfied*	*Somewhat dissatisfied*	*Completely dissatisfied*

This is a non-comparative scale since it deals with a single concept (the brand of a detergent). On the other hand, a comparative scale asks a respondent to rate a concept. For example, you may ask:

Which one of the following brands of detergent do you prefer?

Brand-X *Brand-Y*

In this example you are comparing one brand of detergent with another brand. Therefore, in many situations, comparative scaling presents 'the ideal situation' as a reference for comparison with the actual situation.

3) **Type of Category Labels:** We have discussed different types of category labels used in constructing measurement scales such as verbal categories and numeric categories. Many researchers use verbal categories since they believe that these categories are understood well by the respondents. The maturity and the education level of the respondents influences this decision.

4) **Number of Categories:** While there is no single, optimal number of categories, traditional guidelines suggest that there should be between five and nine categories. Also, if a neutral or indifferent scale response is possible for at least some of the respondents, an odd number of categories should be used. However, the researcher must determine the number of meaningful positions that are best suited for a specific problem.

5) **Balanced versus Unbalanced Scale:** In general, the scale should be balanced to obtain objective data.

6) **Forced versus No forced Categories:** In situations where the respondents are expected to have no opinion, the accuracy of data may be improved by a non forced scale that provides a 'no opinion' category

Check Your Progress-B

1) Explain the difference between comparative and non-comparative scaling techniques with examples.

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

2) Compare the Semantic Differential Scale and Stapel Scale.

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

3) What is construct validity, and how can it be established through convergent and discriminant validity?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

4) Explain the concept of reliability in scaling and discuss two methods to assess it.

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

5) Mark the correct answer

   i) Which of the following is a characteristic of comparative scales?

      a) Evaluates one object independently

      b) Requires respondents to directly compare objects

      c) Uses continuous rating scales

      d) Employs itemized rating scales

   ii) What is the primary limitation of the Paired Comparison Scale?

      a) It generates nominal data

      b) It requires comparing a large number of items geometrically

      c) It does not allow direct comparisons

      d) It provides interval-scaled data

   iii) In which scaling technique are bipolar adjectives used at endpoints?

      a) Likert Scale

      b) Semantic Differential Scale

      c) Stapel Scale

      d) Q-Sort Scale

6.8 LET US SUM UP

This unit outlines fundamental principles and applications essential for developing research instruments and analysing data in social and behavioural sciences. The unit begins by emphasizing the importance of measurement and scaling in quantifying both tangible and abstract attributes, such as customer satisfaction or brand loyalty. Measurement involves assigning numbers to characteristics according to rules, facilitating statistical analysis and communication. Scaling extends measurement by using numerical or semantic rules to represent attitudes, opinions, or feelings.

The unit categorizes measurement scales into four levels: nominal, ordinal, interval, and ratio. Nominal scales classify items into categories without implying order, such as gender or income groups. Ordinal scales rank items based on characteristics but do not quantify the difference between ranks, as seen in customer preferences. Interval scales enable meaningful comparisons by ensuring equal intervals between scale points, though they lack an absolute zero, exemplified by temperature measurements. Ratio scales, with an absolute zero, permit full range data analysis, including ratio comparisons, and are used for metrics like income or duration.

It further explores comparative and non-comparative scaling techniques. Comparative scales, such as paired comparisons, rank order, constant sum, and Q-sort, involve direct comparisons between objects. Non-comparative scales, including continuous and itemized rating scales like Likert, semantic differential, and stapel, measure attitudes or opinions independently of other objects. These techniques are chosen based on research objectives, data requirements, and respondent capabilities.

The unit also discusses reliability and validity, critical for ensuring measurement consistency and accuracy. Reliability includes test-retest, inter-rater, parallel-forms, and internal consistency. Validity covers content, construct (convergent and discriminant), criterion (predictive and concurrent), and face validity, ensuring comprehensive, accurate representation of constructs.

Finally, selecting an appropriate scaling technique depends on problem definition, statistical analysis needs, scale type, category number, and respondent context. The unit underscores that higher measurement levels and balanced, well-constructed scales lead to more robust and interpretable data. This comprehensive coverage provides a framework for effective measurement and scaling in research.

6.9 KEYWORDS

Comparative Scales: Scaling techniques where respondents directly compare one object to another, such as paired comparison, rank order, or constant sum scales.

Construct Validity: Ensures a scale measures the theoretical construct it is intended to represent, verifying alignment with related constructs like convergent and discriminant validity.

Content Validity: Assesses whether a scale comprehensively covers all dimensions and aspects of the concept it aims to measure, ensuring no key elements are omitted.

Criterion Validity: Measures how well a scale's results relate to an external benchmark or standard, such as predictive or concurrent outcomes.

Cronbach's Alpha: A measure of internal consistency reliability, indicating how closely related a set of items are as a group; values above 0.7 are typically acceptable.

Face Validity: The simplest type of validity, referring to whether a scale appears, at a glance, to measure what it is intended to measure.

Interval Scale: A scale with equal distances between values, allowing comparison of differences but lacking a true zero point (e.g., temperature).

Itemized Rating Scale: A scale with predefined categories, allowing respondents to select the category that best represents their evaluation (e.g., Likert or semantic differential scales).

Likert Scale: A widely used itemized rating scale where respondents indicate agreement or disagreement with statements, typically on a five-point scale.

Measurement: The process of assigning numbers or symbols to the characteristics of objects based on pre-specified rules, enabling analysis and interpretation.

Nominal Scale: The simplest measurement scale that classifies data into distinct categories without implying any order (e.g., gender or location).

Non-Comparative Scales: Scaling techniques where respondents evaluate one object independently without comparing it to others (e.g., continuous or itemized scales).

Ordinal Scale: A scale that ranks items based on a characteristic, indicating order but not the magnitude of differences between ranks.

Ratio Scale: The highest level of measurement that includes an absolute zero, allowing for meaningful ratio comparisons (e.g., height or income).

Reliability: The extent to which a scale consistently produces the same results under similar conditions, minimizing random errors.

Semantic Differential Scale: A seven-point scale with bipolar labels (e.g., good-bad) used to assess attitudes or perceptions about objects, brands, or concepts.

6.10 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress Answer A

4 (i) b (ii) d (iii) c

Check your Progress Answer B

5 (i) b (ii) b (iii) b

6.11 TERMINAL QUESTIONS

1) Explain the concept of measurement and scaling in research. How do these concepts assist in achieving research objectives?
2) What are comparative and non-comparative scaling techniques? Discuss their differences and provide examples for each.
3) How should a researcher select an appropriate scaling technique for their study? Discuss the factors influencing this decision with examples.
4) Describe the differences between the Likert Scale, Semantic Differential Scale, and Stapel Scale. Also give examples for each type of scale.
5) Explain the importance of reliability in scaling. Discuss any two methods used to assess reliability with examples.
6) What are the main advantages and limitations of paired comparison and rank order scales in research?
7) How does construct validity differ from criterion validity? Explain with examples.

6.12 FURTHER READINGS

1. Aaker, David A. and George S. Day. (1983) Marketing Research, John Wiley, New York. Bailey, Kenneth D. (1978) Methods of Social Research, The Free Press, New York.
2. Coombs, C.H.(1953) "Theory and Methods of Social Measurement", in Research Methods in the Behavioral Sciences, eds. Feslinger, L. and Ratz, D., Holt, Rinehart and Winston.
3. Donald S. Tull and Gerald S. Albaum. (1973) Survey Research: A Decisional Approach, Index Educational Publishers, New York.
4. Meister, David. (1985) Behavioural Analysis and Measurement Methods, John Wiley, New York.
5. Rodger, Lesile W. (1984) Statistics for Marketing, McGraw-Hill (UK), London.

Note	These questions are helpful to understand this unit. Do efforts for writing the answer of these questions but do not send your answer to university. It is only for your practice.

UNIT 7 QUESTIONNAIRE DESIGN

Structure

7.0 OBJECTIVES

After studying this unit, you will be able to:

- Explain questionnaires, its types, and significance;
- Distinguish between a questionnaire and survey;
- Comprehend the features of a good questionnaire; and

- Learn how the questionnaires are designed and how to frame the different questions corresponding to the information required.

7.1 INTRODUCTION

In a research study—whether exploratory, descriptive, or causal—you need to collect significant amounts of data to answer the research questions. How do you gather this primary data effectively? This is where a questionnaire comes into play. A questionnaire is an instrument used to collect the relevant information from the target participants. It allows for the efficient collection of data from a large number of respondents. A well-designed questionnaire ensures that the data collected is accurate, interpretable, and generalisable, making it an invaluable method for conducting surveys. A well-designed questionnaire leads to the realization of research objectives. It plays a significant role as it facilitates the collection of necessary information in the form of questions.

7.2 MEANING OF QUESTIONNAIRE

A questionnaire is a research instrument that consists of a set of questions or other types of prompts that aims to collect required information from respondents.

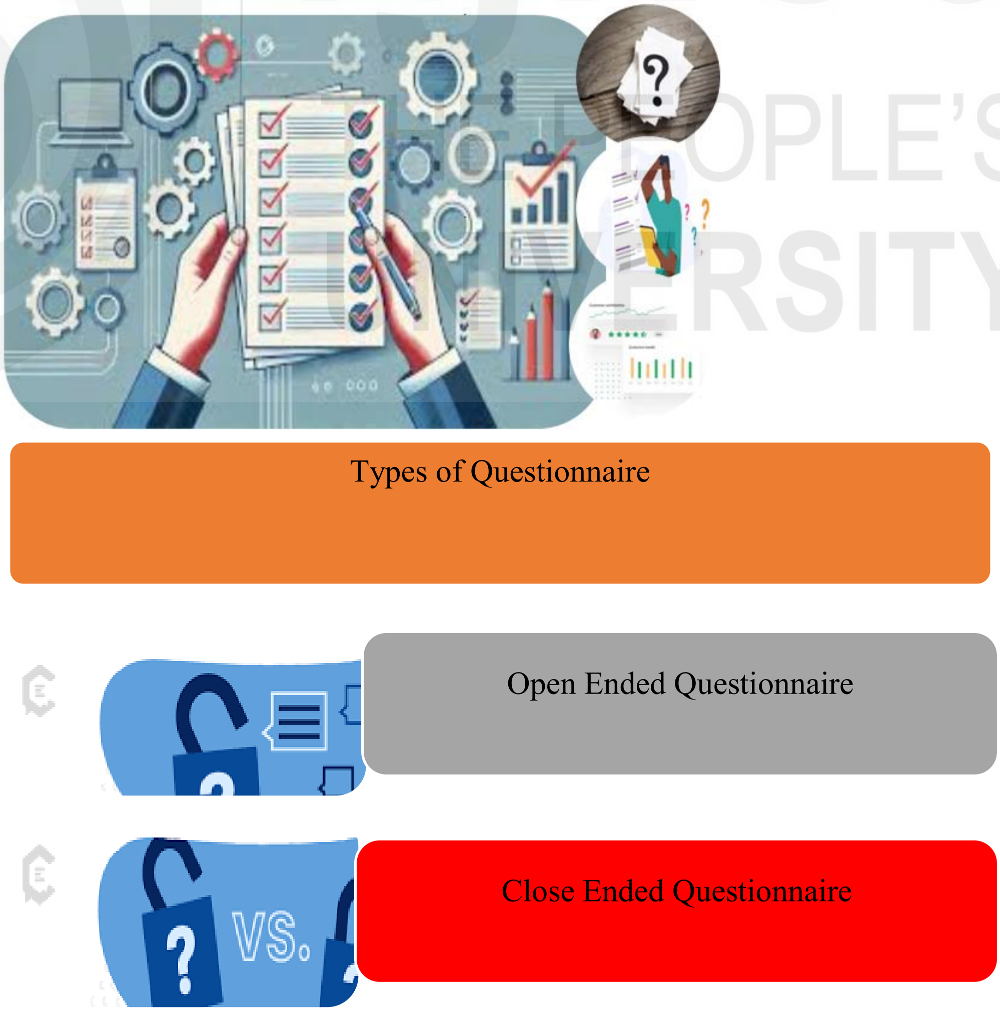


Figure 7.1: Types of Questionnaire

There are two types of questionnaires i.e., open-ended questionnaires and closed-ended questionnaires (refer Figure 7.1). In open-ended questionnaires, there is no limitation on the length of the response, and respondents are free to frame their response. Whereas, in the case of a closed-ended questionnaire, respondents don't have the liberty to formulate their own responses; instead, the researcher creates the questions and provides predetermined responses for every question. Respondents have to select from pre-decided responses.

7.2.1 Questionnaire vs Survey

Questionnaire and survey are the two methods of collecting data from the respondents of the study (refer Figure 7.2). You're probably wondering about the difference between survey and questionnaire. You might be asking about the difference between questionnaire and survey from a business perspective. It's easy to get the two words mixed up. The terms 'survey' and 'questionnaire' are often (wrongly) used interchangeably. This is because people frequently view them as meaning the same thing. Some elucidations on the distinction between questionnaires and surveys may be incorrect. Alternatively, they may be quite confusing. When conducting any form of research, it's important to understand the difference (refer to Table 7.1).

Table 7.1: Difference between Survey and Questionnaire

Survey	Questionnaire
Survey means the collection, recording, and analysis of information **collected** from the respondents.	A questionnaire means a form that contains a set of questions with a view to obtain certain information from the study's respondents.
A survey is a research process used to collect and analyze the information.	The questionnaire is an instrument used to obtain data.
A survey is a time-consuming research process.	A questionnaire consumes less time to acquire the factual information.
The survey is conducted using interview method, questionnaire, telephonic survey, postal or mail out survey, internet-based survey.	The questionnaire is a way of conducting a survey, which is distributed, delivered, or mailed to the respondents.
The survey may include open-ended questions or closed-ended questions, depending on the topic.	The questionnaire includes closed-ended questions only.
The survey is conducted to obtain responses that may either be objective or subjective depending on the questions asked.	The questionnaire is designed to gather objective answers only.
The survey includes not only the collection of information but also observations, measurements, evaluation of collected data, and the researcher's judgment.	The questionnaire includes instructions, questions, and space for responses.

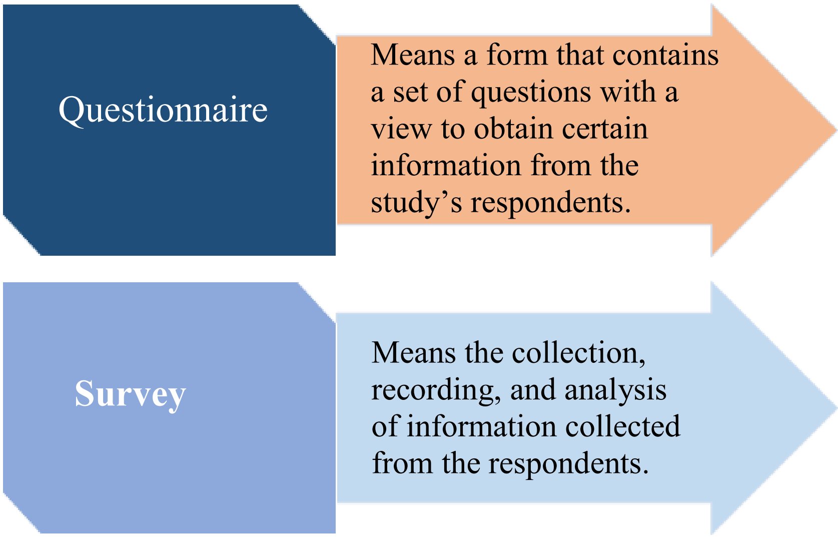


Figure 7.2: Questionnaire and Survey

7.3 SIGNIFICANCE OF QUESTIONNAIRE

A questionnaire is a structured set of questions designed to collect information, opinions, attitudes, or behaviors from respondents. It is one of the most commonly used data collection methods in research. Moreover, questionnaires can be used in various research fields, including social sciences, market research, healthcare, education, and psychology. Their adaptability makes them suitable for investigating diverse research questions.

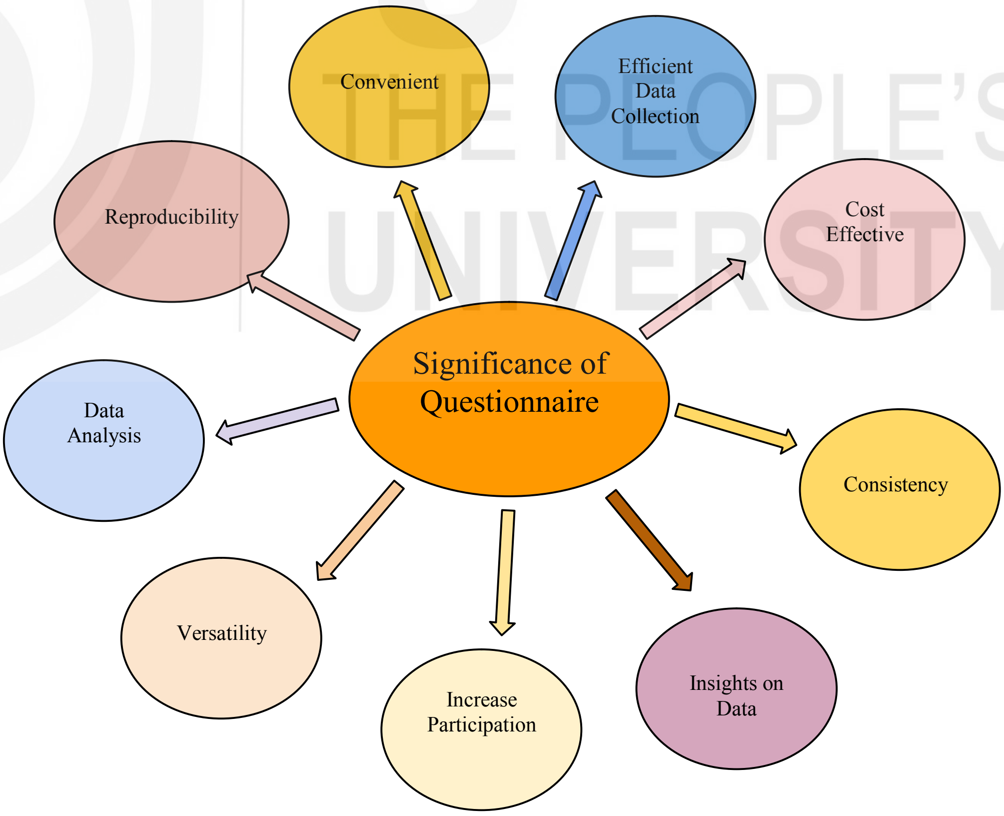


Figure 7.3: Significance/Key Benefits of Questionnaire

"As we know, questionnaires are powerful tools for collecting data from a large number of people. They are especially important in research because they offer several key benefits:

1. **Efficient Data Collection:** Questionnaires can reach a wider audience, allowing researchers to gather data from a large number of respondents at once and quickly, magnifying the reliability and generalisability of the research findings.

2. **Cost-Effective:** Questionnaire administration is comparatively less costly than the other methods of data collection such as interviews or focus groups, especially when administered online.

3. **Consistency:** Questionnaires provide consistency in the data collection by offering the same set of questions to all the target respondents. This makes it easier to compare and analyze the responses.

4. **Insights on Data:** Questionnaire provides quantitative as well as qualitative data for research. Close-ended questions are used to collect quantitative data, which can be easily analyzed statistically that facilitate in identifying trends, patterns, and correlations. While, open-ended questions allow qualitative insights allowing researchers to explore respondent's feelings, thoughts, and experiences.

5. **Increase Participation:** Ethical considerations are followed in the questionnaire. Hence, maintaining confidentiality and anonymity of the respondent increases the number of participants in the study.

6. **Versatility:** Questionnaires can be used in various fields such as education, business, healthcare, and social sciences to gather information on a wide range of topics.

7. **Flexibility in Data Analysis:** Using questionnaires in the study provides robust data analysis as the data collected from questionnaires can be easily organized and analysed using statistical tools and software.

8. **Reproducibility:** A well-structured and well-designed questionnaire can be used again and again in different contexts that enhance the reproducibility of the research findings. It is crucial to validate the results in different contexts and settings.

9. **Convenient:** Self-administered questionnaires offer convenience to the respondents as they can fill it at their pace. It helps in gathering information from people having busy schedules. Also, questionnaires can collect data from a global audience, helping researchers gather data from diverse populations.

7.4 QUESTIONNAIRE DESIGN PROCESS

Designing a questionnaire is dependent on the research objectives and research design. A researcher must be clear about the outcomes of the survey before starting to design a questionnaire.

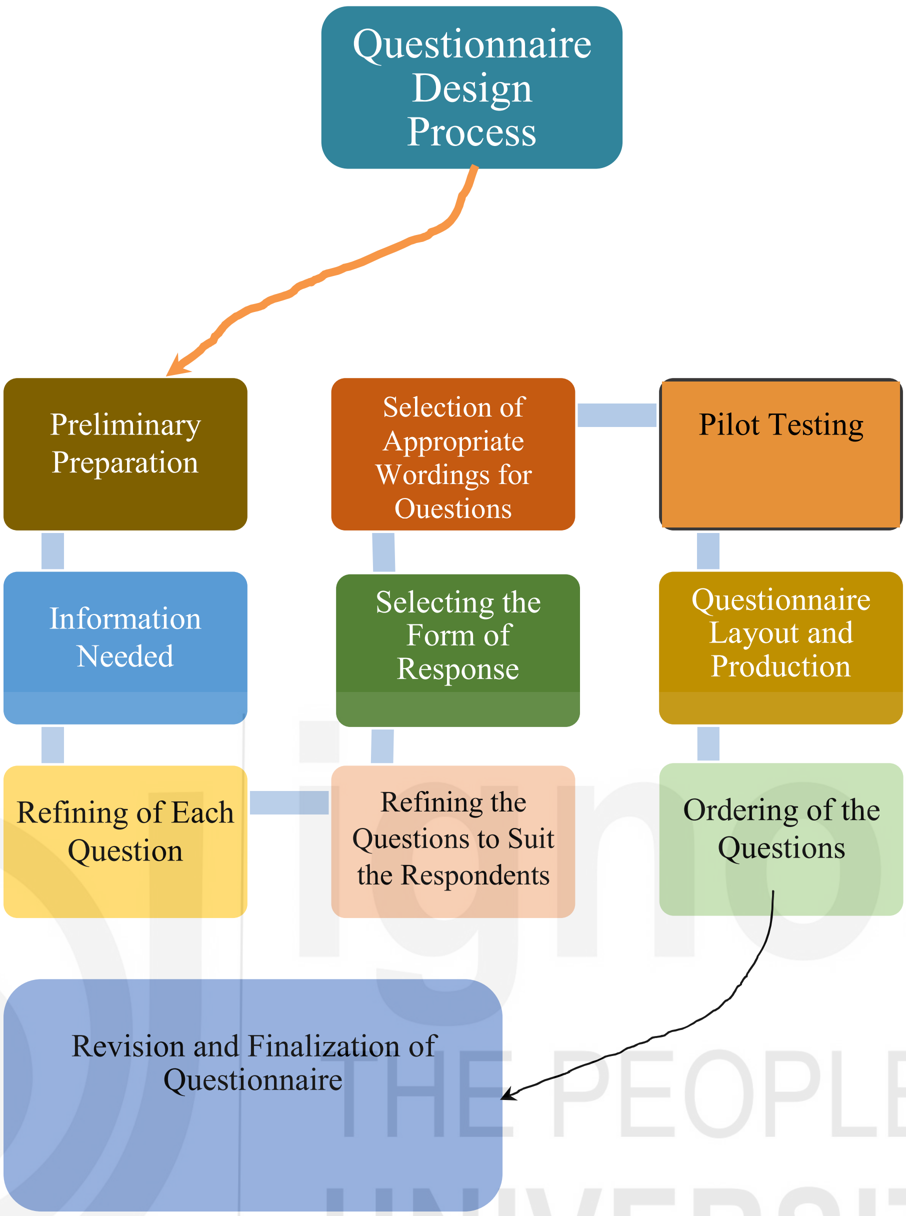


Figure 7.4: Questionnaire Design Process

This process of designing a questionnaire consists of several steps to arrive at the final questionnaire. These steps are discussed below:

1. **Preliminary Preparation:** The questionnaire design process starts with preliminary preparation. Hence, this is the first step of questionnaire designing. It involves the collection of all the relevant information needed for the preparation of a questionnaire. Since the questions may be based on the target respondents, the target population should be well defined. Also, having knowledge about the target population helps in the kind and wording of the questions in the instrument.

2. **Information Needed:** In this step, all the required information to be collected using the survey must be clearly presented by converting into a set of questions. This step ensures that the questions presented in the questionnaire take all the required information for the study. This step must be conducted in a comprehensive manner so that no piece of information is left out.

3. **Refining of Each Question:** Each question in the questionnaire is required to gather some piece of information. Hence, if any of the

questions do not capture the relevant information, then such a question must be removed, and if the researcher finds that the relevant information needed for the study does not have any questions, then those specific questions should be added. This refinement presents a set of questions that will sufficiently collect all the required information.

4. **Refining the Questions to Suit the Respondents:** To ensure that the respondents are capable and willing to respond to the questions, refining the questions is a crucial step. The respondent's ability to answer a question may be affected by many factors, such as lack of adequate knowledge of the research subject, not being able to recall the response easily, difficulty in articulating an answer due to the complexity of the question, the respondent perceiving a question to be out of context, or hesitant to provide sensitive information, etc. Filter questions should be added to ensure the adequacy of knowledge of the respondents. Questions in the questionnaire must aid the memory of the respondents. A list of responses may be provided to remove the complexity of the questions. The questions that require sensitive information are usually put at the end of the questionnaire.

5. **Selecting the Form of Response:** Questions can be classified as structured and unstructured questions based on the form of response. Unstructured questions or open-ended questions are used to capture the general opinion of the respondents, and there is no restriction on the length of the response. While structured questions or closed-ended questions provide respondents with a set of pre-decided response alternatives and assist them in responding to the questions in an easy way. The structured questions are further classified on the basis of the type and number of alternatives provided, such as dichotomous questions, multiple-choice questions, and scale questions. The detailed information on these has been provided later in the unit.

6. **Selection of Appropriate Wording for Questions:** To ensure that respondents are able to comprehend the questions easily, these must be worded in an easy-to-understand way. It is said that answers are as good as the questions. Hence, the researcher must use simple and unambiguous words. The researcher should avoid making implicit assumptions, avoid double-barreled questions, and avoid generalizations as well.

7. **Ordering of the Questions:** An improper order of the questions may discourage the respondents, and that may not respond to the question. Hence, the order of the questions is crucial in generating interest in the target respondents. A questionnaire must start with the opening statements describing the purpose of the instrument, to inform the respondents what would be done with the data collected, and to help them decide whether they are the right respondents for the study. To ensure the familiarity of the respondents with the subject matter and, if not, to filter them out, filter questions may be used at the initial stage of the questionnaire. Initially, easy and interesting questions should be asked by the investigator, setting the context and developing interest in the respondents. Questions that are complex, discomforting, and

sensitive should be at the end of the questionnaire, as initially respondents must be focused on providing the critical information for the research. Identification questions such as respondents' names and contact numbers should be kept at last. Also, a choice must be given to them, as most of the respondents are not willing to disclose their identity. If identity information is crucial for the research, then the purpose must be clearly stated for collecting the identity formation, and if not crucial, then these questions should not be incorporated in the questionnaire.

8. **Questionnaire Layout and Production:** To make a questionnaire convenient to respond to and easy to understand, a questionnaire should be as short as possible, and crowding of questions must be avoided. A questionnaire can be divided into sections focusing on specific areas of study. Color coding in sections may be used to help with easy reference. The questions can be numbered for easy coding of the data collected, and directions for the answers to the questions should be provided clearly and close to the questions. When the final questionnaire is to be administered, the respondent's care must be taken regarding the quality of the questionnaire. Photocopies of the questionnaire should be avoided, and if there are several pages in the questionnaire, a booklet must be prepared. The quality of the paper enhances the professional appearance of the questionnaire.

9. **Pilot Testing:** Pilot testing is crucial for identifying potential issues in the questionnaire regarding content, sequence, difficulty, wording, and instructions of all questions, pilot testing is crucial. It helps in understanding the problems the respondents are facing in answering the questions. In this step, the questionnaire is administered to a small sample of the target respondents, and from the data collected, ambiguity as well as various interpretations by the respondents may be checked, data can also be examined whether the questions are providing the required information. Questionnaire form and layout should also be checked. Hence, inadequate and irrelevant questions can be determined and removed.

10. **Revision and Finalisation of Questionnaire:** All the observations from the pilot testing must be incorporated into the questionnaire so as to ensure that each question serves the purpose of the study, ease of administering the questions, and the flow of the questionnaire. While pilot testing, if a researcher found no changes required in the questionnaire, then the same instrument can be administered for the field survey, and data collected during the pilot study can also be incorporated in the data collected from the field survey.

7.5 PRE-CONSTRUCTION PHASE

The pre-construction phase of instrument/questionnaire designing is mainly focused on planning and conceptualisation. This phase requires defining the research objectives, reviewing relevant literature, deciding on the data collection methods, and identifying the target audience. The stakeholder involvement is vital at this stage, with input from subject matter, researchers, potential respondents, and experts being essential to make sure that the

objectives of the questionnaire are clear and aligned with the research objectives. Ethical issues like consent and confidentiality are also planned during this stage. The outcome of this phase results in a detailed plan and outline that assists as a blueprint for the construction of the actual questionnaire.

7.5.1 Resources required in Designing Questionnaire

Questionnaire design requires mindful planning, various factors, and several key resources. The following are the resources required in designing a questionnaire:

1. **Human Resources:** Human resources involve subject matter experts, researcher(s) who conduct study, data analysts, pilot respondents, and interviewers.

2. **Material Resources:** These resources include survey platforms or software that help in creating and distributing the questionnaire online; examples are Google Forms, SurveyMonkey, Typeform, or Qualtrics. If the survey is being conducted offline, printed materials like pens, papers, and clipboards etc. are used.

3. **Technical Resources:** These resources include tools for survey design and data collection, data analysis software like SPSS, MS Excel, R for analysing quantitative data, and NVivo for analysing qualitative responses. Digital accessibility tools are used to ensure the questionnaire accessibility to individuals with disabilities if distributed online.

4. **Financial Resources:** The researcher should make a budget for tools or software to be used in the study. Sophisticated survey design and analysis tools require subscription fees that must be considered into the budget. These resources also include funds for researchers and experts, and incentives for respondents.

5. **Ethical and Legal Resources:** Resources like institutional standards or guidelines from an ethics review board to make sure the questionnaire respects respondents' rights and privacy. Informed consent forms for obtaining consent from the respondents, specially while collecting personal and sensitive information.

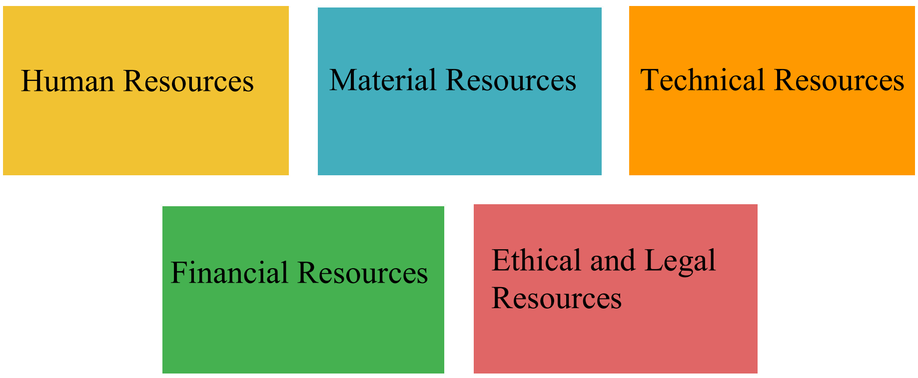


Figure 7.5: Resources required in Designing Questionnaire

7.5.2 Characteristics of Respondents

Understanding target respondents' characteristics is important for developing the questionnaire according to their needs and ensuring valid and reliable data collection.

1. **Demographics:** Age, gender, education level, and occupation are the demographic characteristics of the respondents. People's thinking, perspectives, level of understanding, language, and educational background vary. To develop a questionnaire suitable to respondents and aligned with study's objectives, demographic characteristics must be considered.
2. **Socioeconomic Characteristics:** Respondents' residential status and socioeconomic factors can influence respondents' perspectives, viewpoints, and opinions. Hence, considering these factors is important while selecting the questions for inclusion in the questionnaire.
3. **Cultural and Regional Background:** The words can have different meanings in different cultures. Hence, make sure that questions are culturally appropriate. If your target population consists of respondents from different regions then make sure that questionnaire is translated in their local languages.
4. **Behavioural and Psychographic Characteristics:** While designing questions, a researcher must consider respondents' interests, values, and opinions. A respondent's lifestyle (e.g., active, traveler, sedentary, frequent) may influence their behavior and choices.
5. **Technology literacy and Accessibility:** If the survey is administered online, ensure that respondents have access to the necessary devices and internet. Also, consider the technical literacy of the target respondents is important.
6. **Familiarity with the Topic**: Respondents' prior experience or familiarity with the subject matter may affect their responses. If they do not know the subject matter, they may be reluctant to respond.
7. **Geographic Location:** Geographic location may impact the accessibility to resources, services, and information, affecting the way respondents answer certain questions. In international studies, regional differences may impact responses due to varying economic, political, and cultural contexts.

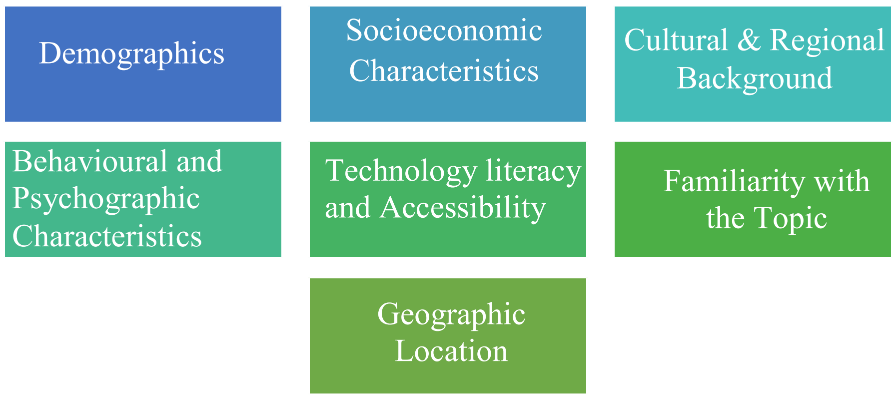


Figure 7.6: Characteristics of Respondents

7.5.3 Selecting an Appropriate Survey Technique

Selection of an appropriate survey technique depends on various factors such as the nature of the research, the characteristics of the target population, the available resources, and the type of data needed. Below are the key steps and considerations when selecting the best survey technique for your research.

1. **Self-administered Questionnaires:** In this survey method, respondents read and respond to the questions themselves without the presence of the interviewer. It is used when the sample size is large, respondents are geographically dispersed, and target respondents are tech savvy. This method is cost-effective, easy to administer, easily scalable, data is easy to collect and analyse.

2. **Face to Face Interviews:** This is the principal method of collecting information. A face-to-face interview is a data collection method where an interviewer and respondent engage in a direct, personal interaction. This method involves the interviewer asking questions and recording the respondent's answers in real-time. Face-to-face interviews can be conducted in various settings, such as homes, workplaces, public spaces, or research facilities. This method ensures a high participation compared to other methods, detailed information, and flexibility.

3. **Telephonic Interviews:** To save time and resources, the researcher might reach out to participants via telephone and pose questions to gather data. This approach is efficient in saving both time and effort; however, it limits the sample to those individuals who have telephone access at home or work.

4. **Internet Survey:** Since the mid-1990s, the internet has emerged as a practical and widely-used method for distributing surveys. There are two primary approaches for utilising the internet as a tool for surveys: via email and web pages.

   i) **E-mail surveys:** E-mail questionnaires are the most economical method among various survey research techniques. They can be categorized into three types. The first type consists of straightforward text questions included within an email. This basic approach allows respondents to edit the email message directly to provide their answers. The second type involves a formatted questionnaire that is sent as an email attachment. The third type is an interactive questionnaire, which can also be sent as an attachment and is designed to be completed and returned via email.

   ii) **Web Page based surveys:** Respondents participating in these surveys need to access a specific web page (URL). On this page, they will have a questionnaire. Usually, participants are recruited for these online questionnaires in one of the following ways:

      - **Pop-up questionnaires:** When someone visits the webpage, a survey may appear, or the visitor may be prompted to go to the URL (web address) to complete a questionnaire This approach to gathering participants relies entirely on individuals randomly

accessing the site and subsequently consenting to participate in the survey.

- **Advertising on other sites:** Various websites might promote a survey. This approach can draw in individuals from many different platforms, resulting in a high volume of responses; however, it can be challenging to identify what demographic such a sample actually reflects.

- **Email invitations:** These may be distributed to individuals with accessible e-mail addresses. Researchers might also utilize a combination of e-mail and web-based surveys. This should be noted that the limitations of this method are largely similar to those associated with telephone interviews.

5. **Structured Observation:** A structured observation gathers data through visual means and is intended to help the observer concentrate on particular actions or features. For instance, if two individuals visit a school and are instructed to count and note the number of computers they observe, check for the existence or lack of air conditioning, and measure the room's dimensions in square feet, they would be engaged in a structured observation. This method is seldom employed in survey research and is more commonly utilized in participant observation.

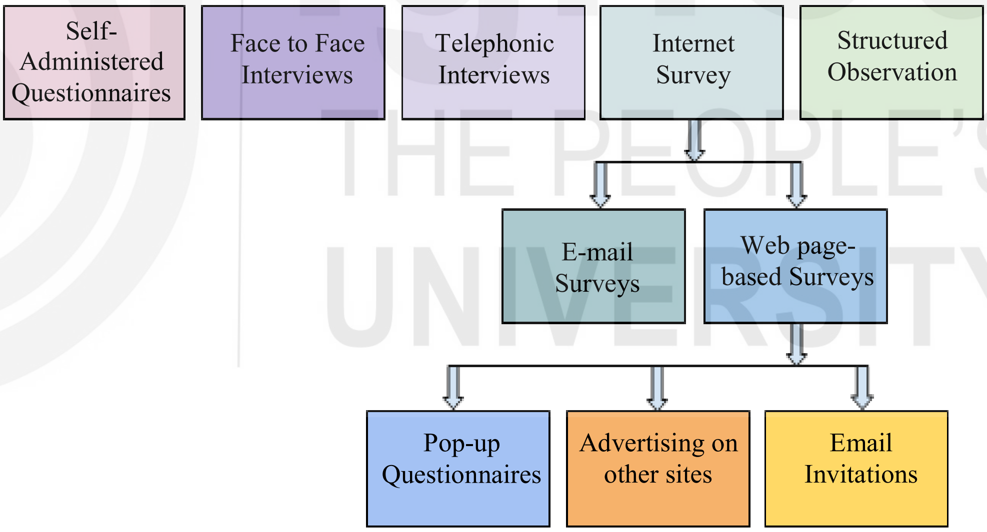


Figure 7.7: Survey Techniques

7.6 CONSTRUCTION PHASE

In contrast to the pre-construction phase of questionnaire design, the construction phase is focused on the practical development of the questionnaire. This step involves selecting appropriate question types (e.g., closed-ended, open-ended), writing specific questions, and designing the layout. Pilot testing is essential during this stage, permitting researchers to test and refine on the basis of feedback received. The emphasis here is on creating relevant and clear, unbiased questions that effectively collect the required data. The construction phase results in a finalized questionnaire that

is prepared for administration, complete with required instructions, and implements all the ethical considerations. Both phases are essential to formulating a reliable and valid questionnaire, with the pre-construction stage laying the foundation and the construction phase bringing the questionnaire to completion.

7.6.1 Questionnaire Format

The questionnaire format refers to the overall structure, design, and organization of the questionnaire, comprising the type of questions, methods for response collection, and grouping and logical sequence of questions. A researcher should decide the format of the questionnaire. The key aspects of questionnaire format are question types, response formats, sectioning or grouping of questions, sequence of questions, instructions, and conditional logic. Only those questions that are relevant for the study are to be preferred. The main objective is to get accurate responses. Questions should be arranged in a logical sequence. Close-ended questions commonly use boxes, brackets, tick marks, or crosses to facilitate quick responses. However, the researcher needs to use one type to maintain uniformity and consistency. There are two types of questionnaires, i.e., structured questionnaires and unstructured questionnaires.

7.6.1.1 Structured Questions

Structured questions refer to questions that are arranged and defined in advance. Structured questions are used in the questionnaire to enhance the consistency of the wording when the study is being conducted at different places and by different interviewers, which ensures that every participant is asked the same question. *For instance, the question "Do you live in India?" can be interpreted differently as "Are you a citizen of India?" by some participants, though it seems that both questions are covering the same information. An individual who is not a citizen of India by currently living in India may respond "yes" to the first question and "no" to the second question.* To increase the reliability of the study and with the intention of asking the question, it is suggested to ask questions either in first form or second form and keep the wording of the questions exactly the same when asked by different interviewers. Structuring responses in a questionnaire also attains consistency. It makes interpretation of responses easier. Types of structured questions include (i) dichotomous questions, (ii) filter or contingency questions, and (iii) level of measurement-based questions.

7.6.1.2 Unstructured Questions

The unstructured questions are those that allow respondents to answer the question in their own words. In other words, these are open-ended questions that have no restriction on the response. Unstructured questions are commonly employed during interviews, where the researcher may not have a predefined set of questions and the sequence of inquiries can be influenced by respondents' answers. This approach often takes place in a relaxed environment. Unstructured questions collect detailed, in-depth, and qualitative information for the study, as they allow respondents to include details and opinions as well. To gather sufficient and necessary information,

the researcher must be attentive, engage in active listening, encourage responses verbally, seek clarifications when needed, and show genuine empathy toward the interviewee.

Unstructured answers suffer from several problems in a comprehensive marketing research study. Unstructured questions can also make respondents take longer to respond, which lowers response rates and produces results of lower quality. Unstructured questions may be interpreted in different ways as well. Due to a lack of predetermined responses, unstructured questions make it more difficult to evaluate data and find trends, patterns, and insights that can guide decision-making.

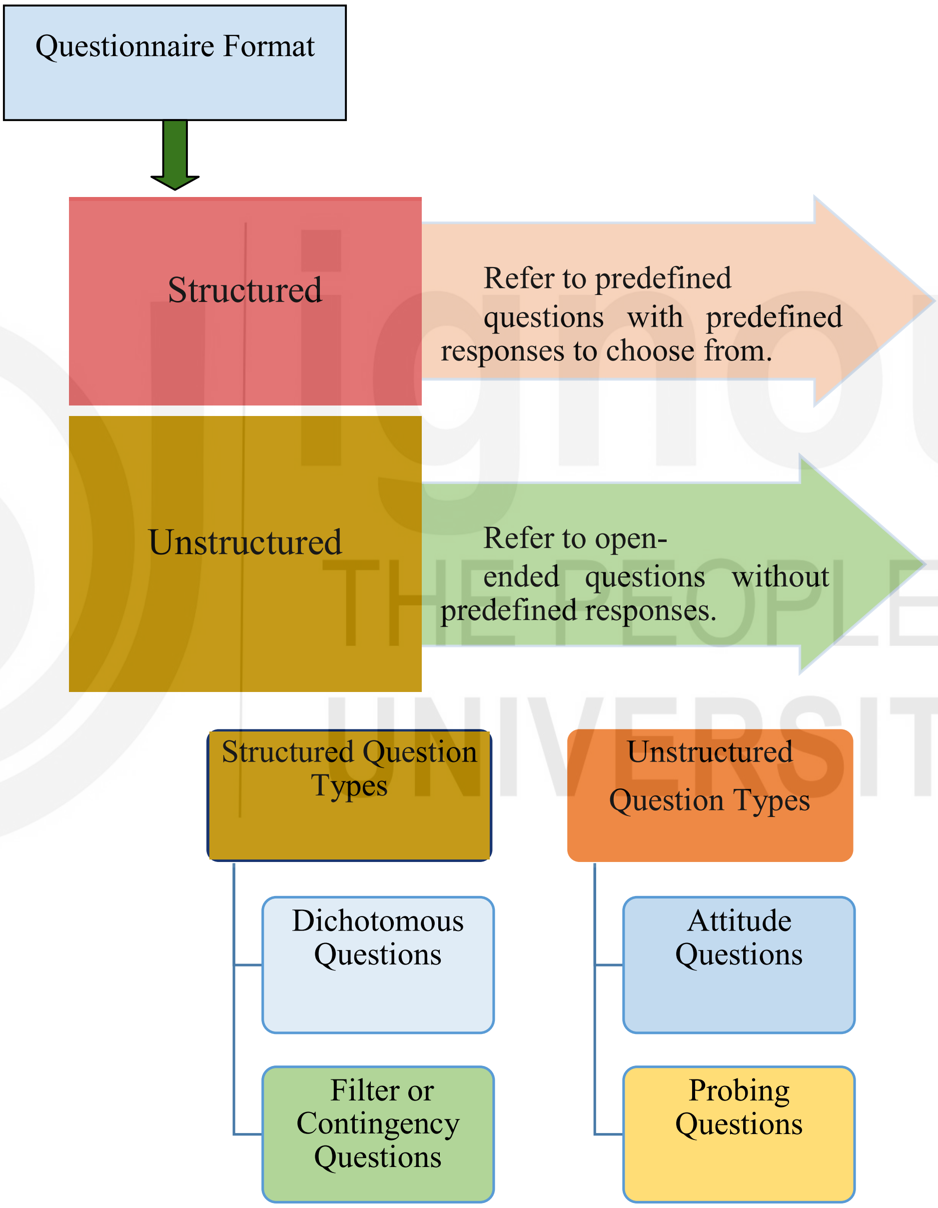


Figure 7.8: Questionnaire Format

7.6.1.3 Structured Vs. Unstructured Questions

Basis	Structured Questions	Unstructured Questions
Definition	Structured questions refer to predefined questions with predefined responses to choose from.	Unstructured questions refer to open-ended questions without predefined responses.
Purpose	Used to collect specific and uniform information or data.	Used to collect insights, experience, and opinions of the respondents.
Format	Closed-ended questions, e.g., True/False, Yes/No, Multiple-choice questions	Open-ended questions, i.e., interview questions, short-answer, long-answer, essays
Flexibility	It follows a strict format and offers no flexibility as respondents have to choose from given options only.	It offers high flexibility as it allows respondents to answer in their own words.
Response Type	Quantitative	Qualitative
Applicability	Surveys, polls, standardised tests.	Interviews, case studies, and focus groups.
Data Consistency	High consistency as it ensures uniformity in the response.	Low consistency as responses can differ widely.
Response Rate	High	Low
Suitability	Quantitative research	Qualitative research
Analysis	Collected data is easy to analyze statistically.	Collected data requires qualitative analysis and can be more complex.
Example	"What is your age?"	"Describe your experience with the product."

7.6.2 Open-ended Questions

Open-ended questions refer to those questions that allow respondents to provide detailed and unrestricted responses in their own words. These questions require detailed and elaborated answers to the questions. Open-ended questions are valuable for gathering qualitative data and in-depth insights, as such questions encourage target respondents to explain, reflect, narrate, or describe. Open-ended questions are a part of an unstructured questionnaire and are designed to encourage conversation, explore the respondents' thoughts, and provide deeper insights. These questions are used when alternatives have not emerged and research is conducted at exploratory stage. Generally, these kinds of questions start with the words such as "How," "What," or "Why". Examples of open-ended questions are:

1. How do you manage work from home?
2. What motivates you to purchase our new product?
3. What are your biggest strengths and why?
4. What did you think of today's workshop?

7.6.3 Close-Ended Questions

Close-ended questions refer to those questions that consist of a set of pre-decided responses. The respondents cannot respond in their own words; instead, they have to choose from available options. They can answer with a single word such as "yes," or "no," or by selecting from the available alternatives. Close-ended questions are used to gather quantitative data, and these are part of a structured questionnaire. These are useful in situations when straightforward and specific information that can be easily quantified and analysed is required. Examples of closed-ended questions are:

1. What is your age in years?

   0 - 18 ____

   19 - 30 ____

   31 - 45 ____

   46 + ____

2. "Airline food service to me is"

   Extremely important____

   Very important____

   Somewhat important____

   Not very important____

   Not at all important____

3. How likely are you to recommend our product to others?

   Very likely____

   Somewhat likely____

   Neutral ____

   Somewhat unlikely____

   Very unlikely____

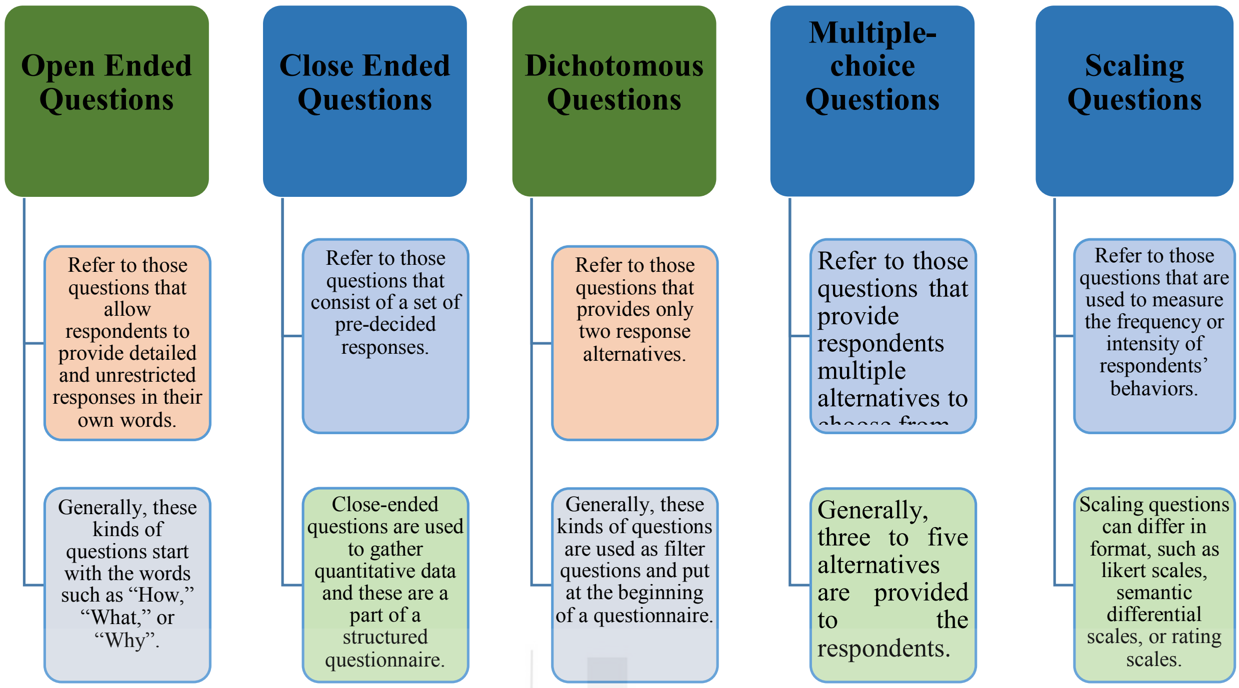


Figure 7.9: Types of Questions

7.6.4 Dichotomous Questions

A dichotomous question is a kind of question that provides only two response alternatives. Generally, these questions are used as filter questions and put at the beginning of a questionnaire. These alternatives are in the form, such as "Yes" or "No," "Agree" or "Disagree," "True" or "False," "Fair" or "Unfair," or any other pairs of mutually exclusive responses. Dichotomous questions are used when the target respondents must strictly fall into either category. Examples of dichotomous questions are:

1. Do you think watching TV can be helpful for children?

   Yes____ No____

2. Product quality improves customer satisfaction?

   True____ False____

7.6.5 Multiple Choice Questions

Multiple-choice questions are a type of question that provides respondents with multiple alternatives to choose from. Generally, three to five alternatives are provided to the respondents. Respondents may choose one or more alternatives. These questions provide more information than dichotomous questions. The above dichotomous questions may be presented in multiple-choice question form.

1. Do you think watching TV can be helpful for children?

   Definitely____

   Definitely not____

   Probably____

   Don't know____

2. Which three features of the product improve customer satisfaction?

   Quality____

   Functionality____

   Ease of use____

   Design____

   Build durability____

7.6.6 Scaling Questions

Scaling questions are a type of question that is used to measure the frequency or intensity of respondents' beliefs, attitudes, feelings, or behaviours. These are very commonly used in surveys. Scaling questions are designed in a way that the responses can be easily and quickly analysed. Scaling questions can differ in format, such as Likert scales, semantic differential scales, or rating scales, and they are crucial in assessing the strength of respondents' sentiments or views about a subject. Examples of scaling questions are:

1. "What type of transportation do you use often?"

   Local Train____

   Bus____

   Metro____

   Car____

   Bicycle____

2. I feel confident using new technology.

   Strongly disagree____

   Disagree____

   Neutral____

   Agree____

   Strongly agree____

Check Your Progress A:

1) Briefly explain the term questionnaire.

   ……………………………………………………………………………

   ……………………………………………………………………………

   ……………………………………………………………………………

   ……………………………………………………………………………

   ……………………………………………………………………………

2) What is a survey?

   ……………………………………………………………………………

   ……………………………………………………………………………

   ……………………………………………………………………………

   ……………………………………………………………………………

3) List the steps of designing a questionnaire.

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

4) Explain the pre-construct phase of a questionnaire.

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

5) Fill in the blanks with the appropriate word(s):

a) A well-designed questionnaire should avoid ______, ______, and ______ questions to minimize bias in answers.

b) Using a ______ scale allows respondents to express varying degrees of agreement or disagreement.

c) The purpose of ______ is to identify and resolve the potential issues in the questionnaire.

d) Using a mix of ______ and ______ questions can help in gathering both quantitative and qualitative data.

e) The ______ of questions should be logical to ensure a smooth ______ for the respondent.

7.7 WORDING OF QUESTIONS

"The wording of questions is crucial in ensuring that respondents clearly understand what is being asked. It involves translating the intended question content and structure into simple, precise, and easy-to-understand language. Poorly worded questions can lead to two major issues:

- **Item Non-Response:** If a question is confusing or unclear, respondents may skip it altogether, leading to missing data, which complicates data analysis and reduces the overall quality of the research.

- **Response Error:** If respondents misinterpret a question due to unclear wording, they may provide incorrect answers. This results in response errors, which can skew the data and lead to biased conclusions.

  To avoid these issues, it's essential that both the respondent and the researcher interpret the question in the same way. Misalignment in understanding can result in seriously biased results, undermining the validity of the research. Following are some dos and don'ts for drafting the questions.

1. **Make use of simple words:** Questions should be formulated using very simple and interactional language. A researcher should avoid the use of jargons, difficult words, short forms, and acronyms. For instance, if a study is being conducted to study the causes of job dissatisfaction in a company, it would be improper to use management jargon such as:

   Which factors are responsible for job dissatisfaction in the company? Select one or more options.

   Low salaries____

   Lack of promotion opportunities____

   Poor work environment____

   workload____

   Rather, questions must be framed using simple words such as:

   What are the factors of job dissatisfaction in the company? Select one

   or more options.

   Low salaries____

   Lack of promotion opportunities____

   Poor work environment____

   workload____

2. **Make use of unambiguous words:** An ambiguous question is one that is confusing due to vagueness, not clearly defining the subject, having more than one meaning, and asking for various responses. There are many words having multiple meanings, which should be avoided in formulating questions in the questionnaire. Hence, to comprehend the question clearly by the respondents, a researcher must ensure that the words mean what is intended and there should not be any scope of ambiguity. Example of ambiguous questions can be:

   How often do you shop online?

   Very often____

   Often____

   Seldom____

   Rarely____

   The options can be understood differently by different respondents, such as some may think shop online twice a month very often, some may find shop online twice a week very often, and twice a month seldom. These interpretations depend upon respondents' perceptions. This ambiguity can be removed by proving the frame of reference. The same question can be framed as:

   How often do you shop online?

   Once a week____

Twice a week____

Once a month____

Twice a month____

Twice in a quarter____

Twice in six months____

Providing reference makes the data collected more meaningful and informative. Also, reference facilitates respondents comprehending the question in the same way as the researcher wants them to understand.

3. **Avoid making implicit presumptions:** Questions as well as alternatives in the questionnaire should not involve any implicit assumptions. All the assumptions of the investigator should be clearly intimated in the questions. For instance, a researcher wants to know the preference of the respondent for the type of music. In other words whether the respondent prefers classical music over jazz. The question is: Do you prefer classical music?

   It is assumed by the researcher that the respondent already knows that the preference of classical music over jazz is being asked, which may not be the case. Thus, the question should be asked as:

   Which type of music do you prefer?

   Classical music____ Jazz music____

4. **Avoid the leading questions:** Leading questions are those that lead respondents to answer a specific response to the questions. Researchers should avoid using these questions in the questionnaire, as these lead to biases in the response, which lead to inaccurate results. *An example of a leading question can be, 'Do you like our excellent new product?'* In this example, the question is biased as it leads the respondents to agree that the new product offered by the company is excellent. The question should be worded as:

   How would you rate our new product?

   Excellent____

   Very good____

   Good____

   Average____

   Poor____

5. **Avoid loaded questions:** Loaded questions refer to the questions that push the respondents to a specific answer. Often this can be referred to as a leading question, but the two are slightly different. When emotional wordings are used to make a question biased, it refers to loaded questions. *For instance, an ABC company does not provide quality goods to its customers. What do you think about the ABC company?* The investigator sets a negative tone about the company.

6. **Avoid double-barreled questions:** These questions refer to the questions that try to take multiple pieces of information at once. In these questions, the researcher asks a lot of information through a single question. Hence, these questions should be avoided by the researcher. An example of a double-barreled question is:

   Do you like the quality and price of this product?

   Yes____ No_____

   This question tries to capture the information about quality as well as price of the product. Some respondents may find the quality of the product as good while some find price reasonable or high price. Hence, some respondents prefer quality and others prefer price. A researcher should avoid the inclusion of these kinds of questions. Rather, researcher should split the question such as:

   1. Do you like the quality of the product?

      Yes____ No_____

   2. Do you like the price of the product?

      Yes____ No_____

7. **Avoid generalisations:** A researcher should not ask respondents to make generalisations or estimates, as it leads them to skip such questions. Questions should be in simple language and specific instead of estimates. *For example, a manager is asked, How many calls do you receive in a year?*

   The manager may find it difficult to recall the number of calls he received during a complete year. Rather, he can easily recall the number of calls on a weekly basis, and after calculating the number of calls on a weekly basis, an annual estimation can be made. The question should be as follows:

   How many calls do you receive in a week?

   When crafting questions for a questionnaire, a researcher should ensure clarity, accuracy, and effective communication with the respondents.

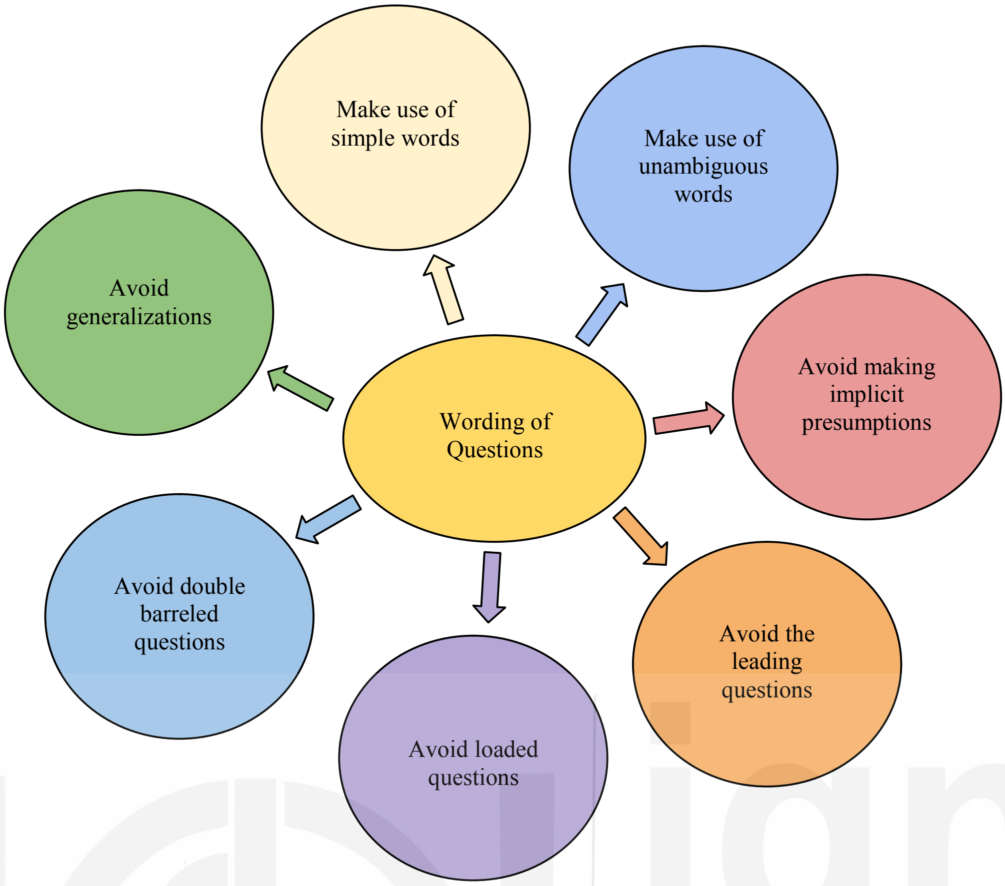


Figure 7.10: Wording of Questions

7.8 SEQUENCING OF QUESTIONS

The sequence of questions in the questionnaire plays a crucial role in generating the interest of the respondents and making them engaged throughout the questionnaire. When the questions are not in proper logical sequence, the respondents may not respond at all. Hence, it is very crucial to arrange the questions in a sequence.

In the beginning, an opening statement describing the purpose of the study must be stated, as the respondents should know the purpose of the study and what would be done with the data collected. Also, initially filter questions can also be used. Filter questions are used when the researcher does not know whether the respondents have knowledge about the subject matter. Filter questions help in filtering out the respondents with insufficient information from the survey.

A questionnaire should consist of easy and simple questions in the beginning to encourage and motivate the respondents to answer the questions. Simple questions facilitate building interest in the respondents.

The opening questions should grab the attention and involve respondents in the interviewing procedure. Complex, sensitive, offending, discomforting, and boring questions should be placed later in the instrument. Questions targeting the collection of sensitive and personal information should be placed at the end of the questionnaire. Also, the questions targeting the collection of classification information such as age, income, gender, etc.

should be placed at the end. One reason for placing classification information at the end is that the collection of relevant information for the study is more important than classification information, as basic questions provide vital information for the research. So, basic questions should be asked before classification questions.

Identification questions such as the respondent's name, contact number, address, etc. should be kept at last in the questionnaire. Also, a choice should be given to the respondents if they wish to provide the identification information. At last, if such information is not important for the study, then there is no compulsion to include these questions in the questionnaire at all.

7.9 RESPONSE CHOICE

A response choice can be defined as the set of pre-decided options provided to the target respondents in a questionnaire. Instead of asking open-ended questions, where respondents can respond in their own words without any restriction on word limits, a response choice limits the answers to specific alternatives available. Standardization of responses makes data collection and its analysis easier to perform. Response choice standardized the responses, which ensures consistency in data collection as every target respondent is provided the same set of responses corresponding to each question. A well-defined response choice facilitates respondents to clearly understand the question, ensuring that collected data is relevant to answer the research questions. Also, it reduces the intellectual load on respondents, which improves the response rate. Every inquiry should include the most comprehensive set of alternatives. This will guarantee that the respondent may select the one that best fits his response. The researcher must brainstorm, examine pertinent published research, consult with specialists, and, if required, lead focus groups with the target respondents to ascertain the best course of action. The researcher can add a category labeled "Other: please specify ________" as one of the alternatives to allow for other options.

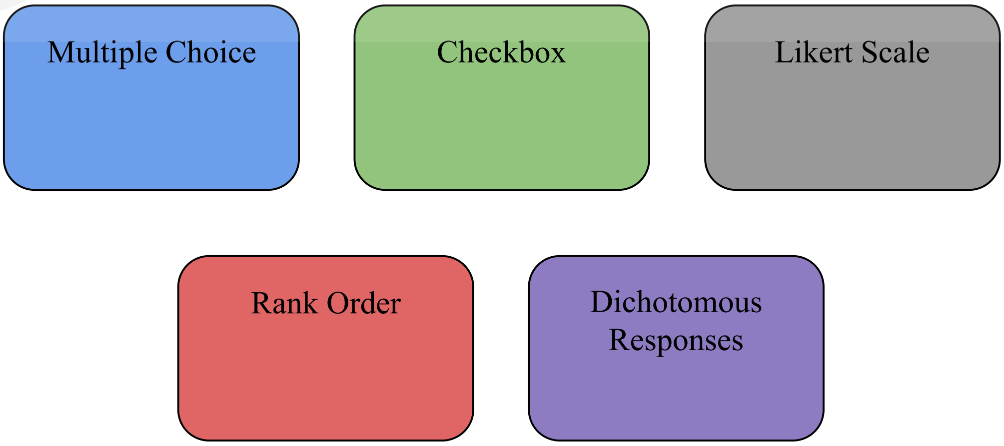


Figure 7.11: Response Choice

Since some respondents might not be aware of the answer to the question, it is crucial to include "don't know" as one of the response options when evaluating factual knowledge. The researcher is "forcing" the responder to guess at a response by not giving them a choice. Some common types of

response choices are given below:

1. **Multiple-choice:** Respondents choose one option from the given set of responses.

   How old are you?

   Under 18 ____

   18 - 24 ____

   25 - 34 ____

   35 - 44 ____

   45 - 54 ____

   55 or Above____

2. **Checkbox (Multiple Selection):** Respondents can choose more than one option from the given set of responses.

   Which of the following activities do you enjoy?

   Reading ____

   Traveling ____

   Watching movies ____

   Cooking ____

   Exercising ____

   Gardening ____

3. **Likert Scale:** It measures the attitude or opinions of the respondents using a scale, usually a 5 or 7-point scale.

   "How satisfied are you with the taste of the food served?

   Very dissatisfied ____

   Dissatisfied ____

   Neutral ____

   Satisfied ____

   Very satisfied ____

4. **Rank Order:** Respondents rank items in order of importance or preference. Rank the following attributes in order of importance when choosing a restaurant

   Food quality ____

   Price ____

   Ambiance ____

   Service ____

5. **Dichotomous Responses:** Respondents have to choose one option from the available two options.

   Did you enjoy the event?

   Yes ____

   No ____

7.10 QUESTIONNAIRE LAYOUT

Questionnaire layout refers to the complete visual design and organization of a questionnaire's elements, focusing on questions, response options, and instructions are spatially organized to improve the readability and respondents' engagement. Physical layout of the questionnaire should be attractive, neat and easy to follow. The questions should be precisely numbered and adequate space should be provided to write the answers. Also, enough spacing between the questions should also be adequately given. This will automatically improve the response rate due to enhanced accuracy and completeness of the questionnaire. Questionnaires, whether mailed or distributed personally, should have an attractive layout to motivate and arouse interest in the respondents to answer the questions. A polite covering letter with a statement such as "thank you for your cooperation" is a must. The questionnaire should make the respondent feel that he/she is a very important component in the research project. In other words, respondents' participation in the study should be appreciated. An example of questionnaire layout is given below in the context of a restaurant survey, which wants to know how it is doing and what needs to be improved.

Restaurant Survey

"Thank you for dining with us!

We value your feedback and would appreciate it if you take a few minutes to complete this survey."

Instructions: "Please answer the following questions based on your recent dining experience."

Name: (Optional)

Date of Visit: ____

Time of Visit: ____

Number of people in your party: ____

"Your Dining Experience"

Question 1: How would you rate the quality of the food you received?

Very Poor ____

Poor ____

Fair ____

Good ____

Excellent ____

Question 2: How would you rate the cleanliness of the restaurant?

Very Poor ____

Poor ____

Fair ____

Good ____

Excellent ____

Question 3: How satisfied were you with the service provided by our staff?

Very Dissatisfied ____

Dissatisfied ____

Neutral ____

Satisfied ____

Very Satisfied ____

"Specific Feedback"

Question 4: Which dishes did you order? (Check all that apply)

Appetizers ____

Main Courses ____

Desserts ____

Beverages ____

Question 5: Were there any menu items you wanted to order but were unavailable?

Yes (please specify):

____

No ____

Question 6: How would you rate the value for money of your meal?

Very Poor ____

Poor ____

Fair ____

Good ____

Excellent ____

"Additional Comments"

Question 7: What did you enjoy the most about your visit?

____

____

Question 8: Do you have any suggestions for improvement?

____

____

"Future Visits"

Question 9: How likely are you to dine with us again?

Extremely Unlikely ____

Unlikely ____

Neutral ____

Likely ____

Extremely Likely ____

Question 10: Would you recommend our restaurant to others?

Very likely____

Somewhat likely____

Neutral ____

Somewhat unlikely____

Very unlikely____

"About You" (Optional)

Question 11: Age Group

Under 18 ____

18 - 24 ____

25 - 34 ____

35 - 44 ____

45 - 54 ____

55 - 64 ____

65 and above ____

Question 12: Gender

Male ____

Female ____

Others/Prefer not to say ____

"Thank you for your time and feedback! We look forward to serving you again."

7.11 FEATURES OF A GOOD QUESTIONNAIRE

A good questionnaire helps you achieve your research goals by providing accurate, complete information and is easy for both the interviewer and respondent to complete.

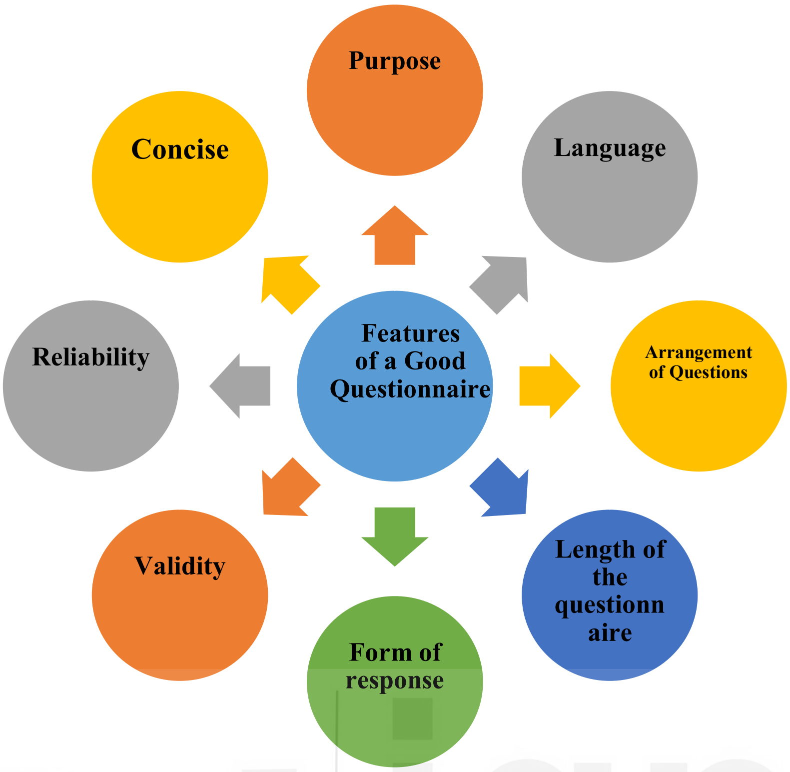


Figure 7.12: Features of a Good Questionnaire

Following are the key features of a good questionnaire:

1. **Purpose:** A good questionnaire should fulfill two purposes. First, it must convert the required information in the form of specific questions, the answers of which provide necessary information to test the hypotheses and provide answers to the research questions. There must be a question corresponding to each information required. The research objectives and the information or data required must precede the construction of the final questionnaire. The second purpose is that the instrument should be designed in such a way that it motivates the target respondents to answer the questions. To ensure the maximum participation of the respondents, it must prove a well-drafted covering letter, which explains the purpose and relevance of the study. It should also assure the respondents that the provided information will be held confidential and used for the research purposes only.

2. **Language:** The questionnaire should consist of questions that are written in a concrete, concise, and straightforward manner to remove any misinterpretation or confusion by the study's respondents. The vocabulary should be easy, syntax must be clear and straightforward. Technical words should be used only if the questionnaire is to be administered on a selected group of professionals in a particular domain. Vague phrases, expressions, proverbs, quotations, and subjective words such as 'good', 'bad', 'fair' etc should be avoided. The language used in drafting the questions should be impartial and not offensive. Researcher's personal perspective should never be represented in the questions, regardless of what he/she may think. In other words, questions should not suggest a certain answer. Keep them neutral, so respondents can answer honestly without feeling influenced.

3. **Arrangement of Questions:** The questions must be arranged in a logical order as it facilitates respondents to maintain a natural flow and keeps target respondents in answering the questions. The initial questions must be easier to answer for the respondents. The question that generates interest should be placed at the beginning of the questionnaire. Generally, the sequence of questions proceeds from simple to complex ones, from general to specific questions, and from those that will create a positive attitude to those that may be sensitive ones.

4. **Length of the questionnaire:** A researcher should keep the questionnaire as short as necessary. If too many questions are asked, the respondents may become tired, and questions at the end could not be well answered. The total number of questions should not be too many to tire or bore the target respondents. Here it becomes important to involve relevant questions only.

5. **Form of response:** The form of answers should be integrated with the form of questions. Although different forms of responses could be used in the same questionnaire. However, it is found that one form is more suitable than other forms for collecting data related to different aspects of the same subject. Dichotomous questions that require responses in either 'yes' or 'no' consist of least bias. Also, these are easy to tabulate, but these types of responses do not collect adequate information from the respondent. In such situations, multiple-choice responses are the desirable form of response.

6. **Validity:** Validity refers to the extent to which a questionnaire measures what is supposed to measure. A valid questionnaire must collect the desired information from the participants. The question should be framed in such a way that the respondents understand how the study is being conducted, i.e., the objective of the study must be clearly underwood by the respondents. In order to do this, the "content expert" should assess the questionnaire during the pilot test (*For example, if the intended respondent is a patient with diabetes, the patient should remark on whether or not he understands the questionnaire*). All the doubts and questions must be solved till the questions are clearly comprehended by the participants.

7. **Reliability:** Reliability refers to the consistency of the questionnaire or instrument with which it measures or assesses what it is intended to measure or assess. In other words, the questionnaire should fetch the same response if the same questions are asked to the participants again and again in a short period of time. This could be achieved by conducting a "test-retest," i.e., administering the same instrument on the participants again and verifying the consistency of the responses. Any inconsistency in the answers may be due to lack of clarity of the questions, which need to be revised and reworded.

8. **Concise:** A brief questionnaire should ask questions intended to address the objectives of the study. Any questions that go beyond the scope of the study have to be eliminated. Generally, researchers frequently "cast the net wider" in an attempt to gather more information regardless of the

significance of the data. This typically occurs when the study objectives have not been adequately considered by the researcher. However, it may pose a risk of asking too many questions, and the questionnaire may run into many pages.

7.12 PRE-TESTING & PILOT TESTING OF A QUESTIONNAIRE

A questionnaire acts as a crucial instrument for gathering data. It should be pre-tested before putting it to actual use. Pre-testing helps in further improving the questionnaire and works like a measuring yardstick, seeking perfection. Once the final questionnaire is printed, then there is no room for corrections and improvement. If the researcher tries to make corrections, it will be expensive as well as difficult. To pretest the questionnaire, it has to be circulated to the small number of respondents to receive useful comments, and the researcher can revise accordingly. Pretest also includes verbal communication with the sample respondents about confusing questions, difficult questions, overlapping categories, etc. In pre-testing, the sample respondents first fill up the questionnaire and then discuss with the researcher the unclear questions. Formal pre-testing is a vital part of the questionnaire design process. It is a learning process for the researcher as well as pre-testing helps the researcher in recording, simplifying, and transforming some of the questions. The process generally involves drafting the questionnaire and discussing it with colleagues, and also circulating it among the small sample of the population for whom the questionnaire is designed. This helps in removing any problem regarding the clarity, understanding of technical or professional terminology, order of questions, etc. Paul Burton (1990) states pretesting may therefore go through a number of iterations, but this is a necessary part of questionnaire design. Pretesting, also called pilot survey, is very much desirable to finalise the questionnaire. Krishan Kumar (1992) has also suggested that a second pretest must be carried out with a revised questionnaire. He further observes that if some difficulties occur, another pretest might be required. Pre-testing also indicates the time required to fill up the questionnaire. It is a practice not to include the pretest sample to the actual population.

Pilot testing is also known as pre-testing. It is very important as it increases the use of any tool that is used to collect the primary data. A pilot test is done by giving the questionnaire to a smaller set of people and the responses are collected and interpreted. These interpretations should be checked with the required answers and objectives to see whether they are meeting the criteria or not.

Check Your Progress B:

1) What is pilot-testing of a questionnaire?

…………………………………………………………………………….

…………………………………………………………………………….

…………………………………………………………………………….

2) What is the response choice?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

3) What is the validity and reliability of a questionnaire?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

4) Define the questionnaire layout?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

7.13 LET US SUM UP

This unit explained the meaning of the questionnaire, its significance, and how a good questionnaire can be designed. To design a good questionnaire, a researcher should ensure that the questions must fulfill the agenda of the research; questions must be written in simple language and logically arranged. Validity and reliability of a questionnaire is a must to draft a good questionnaire. All the questions must fetch a piece of relevant information for the research. A questionnaire can be an unstructured or structured questionnaire. Unstructured questionnaire is used to collect qualitative information and structured questionnaire is used to collect quantitative information. Open-ended questions are those that offer respondents a predefined set of questions whereas close ended questions do not restrict the response and the respondent can draft response in their own words. The response form must be suitable to the form of the questions asked as a researcher cannot use the same form for different kinds of questions. Dichotomous questions consist of least biasness and provide limited information as a respondent is provided only with two alternatives such as “Yes/No” or “True/False” etc. While multiple choice questions provide adequate information by providing more than two options corresponding to each question. Scaling questions are used to measure the respondents’ beliefs, attitudes, feelings, or behaviors using different scales such as Likert scale, semantic differential scales or rating scales. Questionnaire formatting and layout must consider the wording and sequencing of the questions. This will motivate the respondents to answer thereby improving response rate. Before

administering the final questionnaire, pre-testing is crucial as it helps in removing ambiguity of the questions. Expert opinions and respondents' opinions should be implemented in the questionnaire to make it an effective, valid and reliable instrument and realize research objectives.

7.14 KEYWORDS

Double-Barreled Questions: Double-barreled questions refer to those questions that try to take multiple pieces of information at once. In these questions, the researcher asks a lot of information through a single question.

Pilot-Testing: Pilot testing of a questionnaire can be defined as a process of administering a preliminary test of the instrument on a small number of representative samples from the target population to identify and resolve potential issues such as confusing wordings, ambiguous questions, and logical sequencing of the questions.

Questionnaire: A questionnaire is a research instrument that consists of a set of questions or other types of prompts that aims to collect information from respondents.

Questionnaire Layout: Questionnaire layout refers to the complete design and organization of a questionnaire. The physical layout of the questionnaire should be attractive, neat and easy to follow.

Response Choice: A response choice can be defined as the set of predecided options provided to the target respondents in a questionnaire.

Reliability: Reliability refers to the consistency of the questionnaire or instrument with which it measures or assesses what it is intended to measure or assess.

Validity: Validity refers to the extent to which a questionnaire measures what is supposed to measure. A valid questionnaire must collect the desired information from the participants.

7.15 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress-A

5. Fill in the Blanks

a) leading, loaded, double-barreled

b) Likert

c) pilot-testing

d) closed-ended, open-ended

e) order, flow

7.16 TERMINAL QUESTIONS

1) Define a questionnaire. What is the significance of a questionnaire?
2) Distinguish between a questionnaire and a survey.
3) Briefly explain the process of designing a questionnaire.
4) Explain the steps to be followed in the pre-construct and construction phases of a questionnaire.
5) What are the features of a good questionnaire? Briefly explain.
6) Briefly describe the do’s and don’ts for drafting the questions.
7) Describe the sequence of questions in a questionnaire with the help of an example.

7.17 FURTHER READINGS

1. "Research Methodology: Methods and Techniques" by C.R. Kothari
2. "Marketing Research: An Applied Orientation" by Naresh K. Malhotra
3. Fred N. Kerlinger. Foundations of Behavioural Research, Surjeet Publications, Delhi
4. "Designing Surveys: A Guide to Decisions and Procedures" by Johnny Blair, Ronald F. Czaja, and Edward A. Blair
5. "Research Methods for Management" by Deepak Chawla and Neena Sondhi
6. "Questionnaire Design: How to Plan, Structure, and Write Survey Material for Effective Market Research" by Ian Brace

Note	These questions are helpful to understand this unit. Make an effort for writing the answer of these questions but do not send your answer to university. It is only for your practice.

UNIT 8 SAMPLING DESIGN

Structure

8.0 OBJECTIVES

After studying this unit, you will be able to:

- Understand the meaning of sampling;
- Differentiate between population and sample;
- Explain the concept of sampling frame and sampling unit;
- Discuss the types of sampling, including probability and non-probability methods;
- Understand the central limit theorem and its importance in sampling;
- Identify types of sampling errors and strategies to reduce them; and
- Discuss the significance and characteristics of a good sample.

8.1 INTRODUCTION

The preceding unit of this course has emphasized that the process of developing "research design" is the most difficult undertaking that follows the task of identifying the research problem. The research design includes all strategic choices concerning the extent, site, timing, quantity, and methodology of an investigation or research study.

In essence, the research design is the theoretical structure in which research is carried out; it acts as the plan for gathering, quantifying, and examining data.

Therefore, the design includes a thorough strategy that outlines the researcher's procedure, starting from the development of the hypotheses and its practical consequences to the ultimate data analysis. The comprehensive research design consists of the sampling design, observational design, statistical design, and operational design.

This unit is about sampling design, which is a detailed plan for getting a sample from a certain population. It refers to the method or process the researcher would use to choose the topics for the sample. The number of items in the sample, which is also called the sample size, may be specified in the sample design. Sample design is decided on before any data is collected. Many sample designs are available for researchers to pick from. It's easier and more accurate to follow some designs than others. According to the researcher, it is important to choose a sample design that is reliable and fits in a research study. The researcher should choose the sample design based on the study's purpose and other factors. But, before discussing sampling examples and applications, it is required to define sampling concepts and principles, they are discussed below :

1. **Universe & Population:** From a statistical perspective, the term "Universe" refers to the total number of items or units in any field of inquiry, while "population" refers to the number of items about which information is desired. Characteristics and elementary units are the subject of study. Population is the sum of such units. All units in any field of inquiry are universes, and all elementary units (based on one or more characteristics) are populations.

2. **Sampling frame:** A sampling frame is a list of all the elementary units or items in a population from which a sample is drawn. It serves as the source material or device for selecting a sample in a study. Ideally, if the population is finite and current, the sampling frame should match the population exactly. However, in practice, obtaining a complete and up-to-date list of the entire population is often challenging, so the sampling frame may differ from the actual population. Either a researcher created this frame for his study or a population list was used. Telephone directories can be used for citywide opinion surveys. A frame should accurately represent the population.

3. **Sampling Design:** A sampling design is a concrete plan that outlines the process for selecting a sample from the sampling frame. It specifies the researcher's method for choosing sampling units to draw inferences

about the population. The sampling design is determined before data collection begins to ensure the research is methodologically sound. Different types of sampling designs will be discussed later in this unit.

4. **Statistic(s) and parameter(s):** Statistics are sample characteristics, while parameters are population characteristics. Thus, metrics like mean, median, mode, and others are called statistics because they describe a sample's characteristics. However, such measures are called parameters when they describe a population. The population mean (μ) is a parameter, while the sample mean ($\bar{x}$) is a statistic. To estimate a parameter from a statistic is the main goal of sampling analysis.

5. **Sampling error:** When Sample surveys involve a small portion of the population, the data collected may be inaccurate. This inaccuracy is called sampling error or error variance. Sampling errors are errors caused by sampling and usually involve random variations in sample estimates around true population values.

In this section we introduced you to some basic concepts and terminologies which will help you to understand the content presented in this unit, in a better way.

We are now in position to understand the significance of sampling, the same is discussed below in the subsequent section.

8.2 POPULATION AND SAMPLE

Population and sample are two ideas that go together and form the basis of research design. In research, the population is the whole area of interest, and the sample is a smaller group that can be used to gather data. How well the sample represents the whole population affects how valid and trustworthy the research results are. Scholars who want to learn more about the population use sampling to help them draw conclusions that are useful for the whole population. Knowing the difference between a population and a sample is important for conducting good research and drawing correct conclusions.

8.2.1 Population

In research, the population is the whole group of people, things, or events that the researcher wants to learn about. Depending on the study's goals, this could be anyone from all the students in a school to everyone who lives in a country. The researcher set the criteria for the population, which includes everything that meets those criteria. In theory, it gives us the most complete set of data because it includes all the observations that can be made. But it's not always possible to study a whole population because of constraints like time, money, and access. This is where the idea of a sample is very important.

8.2.2 Sample

A sample, on the other hand, is a subset of the population that is chosen for the study. The sample is chosen so that it accurately represents the

population, which means that it has the same characteristics as the population as a whole. Using a sample lets researchers draw conclusions about the whole population without having to collect data from every person or unit in it, which makes the research process faster and easier to handle.

8.2.3 Relationship between Population and Sample

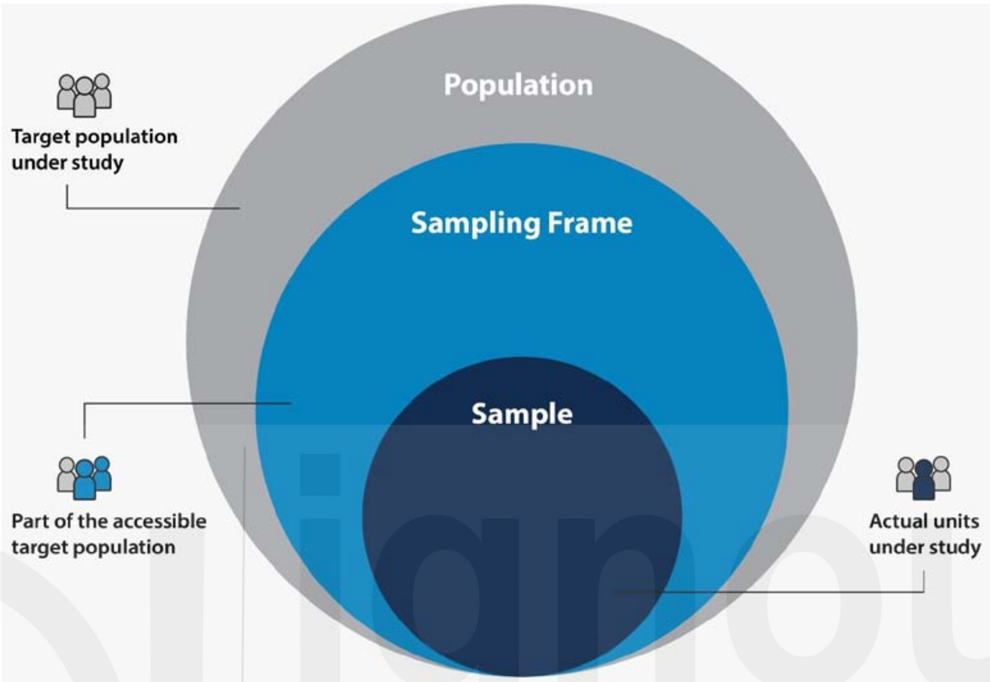


Figure 8.1: Relationship between Population and Sample

The relationship between a population and a sample is very important because it determines how true the research results are for the population as a whole. A sample is a powerful tool for researchers because it can give results that can be applied to the whole population if it is chosen correctly. *For example, if the sample in a survey about consumer behaviour is a good representation of the target population in terms of demographics, then the results can be confidently applied to the whole group.*

On the other hand, mistakes way happen while sampling. There is always spatial error, which is the difference between the sample results and the true values for the whole population. This is true even if the sample is carefully chosen. But when researchers work with samples instead of whole populations, they can't avoid this error, even though they use statistical methods to estimate and lower it. The difference between studying a sample and the whole population in this case shows the trade-off between the two goals.

In real life, a population is the full set of data and the best standard for research, but a sample is a more useful way to collect and analyse data. What makes a sample a good representation of the whole population depends on how it was chosen. Probability sampling methods, like simple random sampling or stratified sampling, are meant to keep bias to a minimum and make sure that the sample is very similar to the population as a whole. Non-probability sampling methods, such as convenience sampling, may be easier

to use, but they have a higher chance of bias, which could make the results less useful in real life.

The decision to study a population or a sample is ultimately determined by the research goals, the resources at hand, and the necessary level of accuracy. Although population study would theoretically yield the most precise and thorough data, sampling provides a pragmatic option that strikes a balance between the need for dependable results and the limitations of research.

Remember, a statistic is a property of a sample, while a parameter is a property of the whole population. Therefore, the average, median, mode, and other numbers gathered from samples are known as statistics because they describe the sample as a whole. Although, these kinds of measurements are called parameters when they describe the traits of a population. A good example of a statistic is the sample mean ($\underline{X}$), while a parameter is the population mean (μ). An estimate of a parameter from a statistic is one of the main goals of sampling analysis.

Now, we will discuss the concept of sampling in the subsequent section

8.3 SAMPLING

Sampling is the selection of part of an aggregate or totality to make a judgement or inference. Thus, it involves studying a portion of a population to learn about it. Most research and surveys make generalisations or draw inferences from samples about population parameters. Researchers usually choose a few items from the universe to study. We assume the sample data will allow them to estimate population parameters. The items selected are called a sample, their selection process is called sample design, and the survey based on the sample is called sample survey. To draw valid and reliable conclusions, the sample should be truly representative of population characteristics without bias.

All research items form a ‘Universe’ or ‘Population.’ A census inquiry lists all items in the ‘population’. Assuming all items are covered, such an inquiry eliminates chance and maximises accuracy. In practice, this may not be true. Any bias in such an inquiry will grow as the number of observations increases. There is no way to check bias or its extent without resurvey or sample checks. This type of investigation is also time, money, and energy intensive. When the field of inquiry is large, this method is difficult to adopt due to resource requirements. This method is sometimes beyond ordinary researchers' reach. The government may be the only institution that can complete the enumeration. Even the government uses this in rare cases like a decade-long population census. Additionally, it is not always possible to examine every item in the population, and studying only a portion of the population can yield accurate results. In such cases, census surveys are useless.

It's important to note, though, that a sample survey doesn't work when the universe is small. When field studies are done in real life, constraints like time and money are always there which mean that only a few items are

chosen from the respondents. To make a miniature cross-section, the people who are asked to fill out the survey should be as typical of the whole population as possible. The people who were chosen are technically called a "sample," and the method used to choose them is called a "sampling technique." This kind of survey is called a "sample survey." To use algebra, let the size of the population be N. If a group of n units (less than N) from this population is chosen according to a rule to study some property of the population as a whole, this group is called a "sample." The researcher needs to make a sample design for his study. This means that he needs to plan how to choose a sample and of what size sample should be.

It is important to remember that in research, a sample is a part of a population that is chosen to be studied. The traits of a sample are very important because they show how well the sample represents the whole population. This, in turn, impacts the reliability and applicability of the research results. The sample's characteristics are talked about in the next section. We will discuss types of sampling later in this unit.

It is to be noted that, the sample design to be used must be decided by the researcher taking into consideration the nature of the inquiry and other related factors.

8.3.1 Characteristics of a good Sample

Synchronizing with the characteristics of a sample, is the key behind the success of any research project. A well-chosen sample can give you accurate information about a population if it is of the right size, was picked at random, is independent, easy to access, can be measured, is stable, precise, and doesn't cost too much. So, the sample is a very important part of the research process because these traits make it likely that the results from the sample can be confidently applied to the whole population.

The key characteristics of a sample, are discussed below:

1. **Representativeness**: A sample effectively represents the attributes of the population from which it is selected. This implies that the sample should include the same proportions of variations prevalent in the population, such as age, gender, income level, or other pertinent characteristics. It is a crucial characteristic because the higher the level of representativeness of a sample, the more researchers can confidently extrapolate their findings to the whole population. Representativeness deficit can result in biased findings that fail to accurately represent the population.

2. **Adequate Size**: In any study, sample size refers to the number of participants or observations. The sample size must be large enough to represent the population accurately. A larger sample size reduces sampling error and improves research accuracy. Given the study's practical constraints, the sample size must be reasonable.

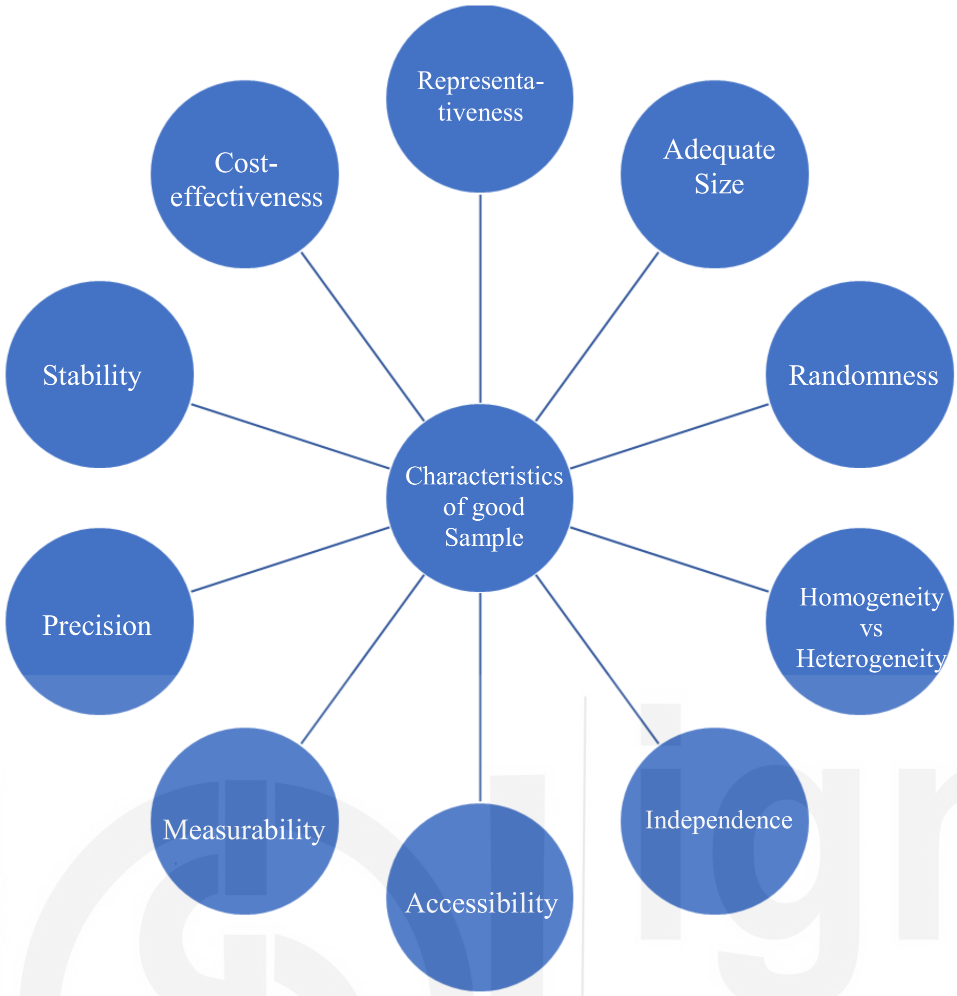


Figure 8.2: Characteristics of a Good Sample

3. **Randomness**: Randomness in sampling refers to the condition where each element of the population has an equal probability of being selected for inclusion in the sample. This attribute is essential for eradicating selection bias. Random sampling methods are crucial as they establish the unbiasedness of the sample and enable the generalization of results to the population. Selection of non-random samples can introduce systematic biases, resulting in erroneous conclusions.

4. **Homogeneity vs. Heterogeneity**: The homogeneity of a sample refers to the degree of similarity among its elements, while heterogeneity denotes the degree of dissimilarity among the elements. This characteristic is significant because the sample should exhibit equality in terms of homogeneity or heterogeneity with the target population. Given the diversity of the population, it is imperative that the sample also exhibits diversity, encompassing all types of individuals. Conversely, if the population is rather homogeneous, a sample that closely resembles the population may just be sufficient.

5. **Independence**: Independence within a sample refers to the property where selecting one element does not alter the selection of another element. Each individual sample unit can be selected independently. The significance of this property lies in its ability to reduce the likelihood of bias during the sample selection process. Thus, the gathered information

from one individual or object will not alter the information obtained from another, thus ensuring the integrity of the research.

6. **Accessibility**: Accessibility is a measure of how easy it is to get in touch with, observe, or study the chosen sample units. This is a necessary property for data collection to work because a sample needs to be easy to get to. This trait can make it easier or harder to do research and changes the sampling method that is used.

7. **Measurability**: Measurability refers to the ability to precisely measure or observe the inherent qualities of a selected sample. Ensuring the measurability of the variables of interest is a crucial attribute for a sample to be valuable in research. If specific attributes of the sample cannot be precisely quantified, it undermines the credibility of the research results.

8. **Precision**: When you use statistics, "precision" means how closely the sample estimates, like the mean or the proportion, match the real population parameters. A very high level of precision is very important because it makes sure that the estimates from a sample are very close to the real values of the population. To come to correct conclusions, precision is very important. Things like the size of the sample and the method used for sampling play a big role in this.

9. **Stability**: Stability, in the context of sample characteristics, refers to the degree of consistency observed among multiple samples obtained from a single population. This attribute is of utmost importance since, if the sampling procedure is replicated, a consistent sample will produce similar results. It is this quality that determines the reliability of the research findings.

10. **Cost-Effectiveness**: Cost-effectiveness refers to the balance reached between the quality of the sample and the resources required to obtain it economically. An essential attribute of a high-quality sample is its ability to provide reliable data without incurring unnecessary costs. The acquisition of a sample necessitates researchers to consider both the financial and temporal costs.

Before proceeding to the subsequent sections of this unit, let's give a glance to the steps involved in the development of sample design, this will help you understand the relevance of other sections of this unit, in advance.

Steps In Sample Design

While developing a sampling design, the researcher must pay attention to the following points:

i) **Select type of universe**: Before creating a sample design, define the Universe, or set of objects to be studied. The universe can be finite or infinite. In a finite universe, the number of items is known, but in an infinite universe, we don't know the number of items. A city's population, a factory's workers, and the like are finite universes, while stars in the sky, radio listeners, dice throwing, etc. are infinite universes.

ii) **Define Source list**: This is the "sampling frame" from which to draw a sample. In a finite universe, it lists all items. Researchers must create source lists if unavailable. A complete, accurate, reliable, and appropriate list is needed. As much population representation as possible in the source list is crucial.

iii) **Identify Sampling unit**: After selecting a sample frame, a sampling unit must be chosen. A sample unit can be a geographical unit (state, district, village, etc.), a construction unit (house, flat, etc.), a social unit (family, club, school, etc.), or an individual. The researcher must choose one or more units for his study.

iv) **Identify Size of sample**: Sample size is the number of items chosen from the universe to form a sample. A researcher faces a major issue. Sample size should not be too big or small. It should be ideal. The best sample is efficient, representative, reliable, and flexible. Researchers must consider precision and estimate confidence when choosing sample size. When population variance is high, a larger sample is needed. Keep in mind that population size limits sample size. When choosing a sample size, research study parameters must be considered. Our sample size is also limited by costs. Budgetary constraints must be considered when choosing a sample size.

v) **Select Parameters of interest i.e. Population parameters**: When choosing a sample design, consider the population parameters of interest. For instance, we may want to estimate the population's proportion of people with a certain trait or know its average. We may also want to estimate important population subgroups. This strongly influences the sample design we accept.

vi) **Monitor Budgetary constraint**: From a practical standpoint, cost considerations significantly influence decisions regarding both the sample size and the sample characteristics. This reality can, in fact, result in the utilisation of a non-probability sample.

vii) **Specify Sampling procedure:** Ultimately, the researcher must decide on the suitable sample type and selection technique. This approach exemplifies the sample design. The researcher is required to select one of four sample designs (elucidated in the subsequent pages) for his study. Clearly, he must choose the design that, given a specific sample size and cost, exhibits a lower sampling error.

After learning about the characteristics of sample and the steps involved in the designing of the sample, it is required to learn about the Sampling & Non-Sampling Errors. Earlier in this unit we had briefly discussed about the sampling errors, now we will discuss the Sampling & Non-Sampling Errors in detail.

8.3.2 Significance of Sampling

Sampling is important in research because it makes it possible to study large groups of people in a quick, accurate, and useful way. Sampling is a reliable way to get information from a small part of a larger population. This

information helps researchers come to meaningful, general conclusions that advance knowledge and help people make decisions in many areas. Its ability to lower costs, improve data quality, and support inferential statistics shows how important it is as a basic tool in research methods.

Collecting samples is an important part of research methods because it lets you study large groups of people without having to look at each individual or unit in the group. Sampling is important for many parts of research because it leads to efficiency, accuracy, and reliability, all of which are necessary for finding meaningful insights.

When doing research, it is not always possible to study a whole population because it takes too much time, money, or complexity with logistics. These problems can be solved by sampling, which lets researchers focus on a smaller, easier-to-handle portion of the population. This not only makes it easier to do research, but it also makes efficient use of resources. *For example, instead of polling millions of people across the whole country, researchers can choose a sample that accurately reflects the characteristics of the larger population.* This lets them get important information faster and for less money.

In addition, sampling improves the quality of the data that is gathered. By focusing on a smaller group, researchers can pay more attention to making sure the data is correct and complete, which is very important for getting accurate results. If you carefully plan and use a sample, you can get results that are just as accurate as those from a study of the whole population. This trustworthiness is increased by reducing any possible biases that might change the results. This makes sure that the conclusions from the sample are true for the whole population.

Sampling is important for more reasons than just being useful and efficient. It is the basis for inferential statistics, which is an important field of research that lets us apply what we learn from a sample to the whole population. Researchers can use data from sampling to make predictions and smart decisions. Some of the techniques they use are hypothesis testing and estimating population parameters. In fields like public health, where accurate and up-to-date data help formulating policy decisions that affect a lot of people, this ability becomes really useful.

Sampling is also very important for research that involves large groups of people, since it would be impossible to get information from every single person. Researchers can still study population trends, ehaviours, and characteristics by choosing a sample that is representative of the whole. This way, they can learn things that can be used for the whole population. This method is not only useful, but also moral, since it makes things easier for participants and values their time and resources.

In short, Sampling is used for many reasons, including:

1. Sampling saves time and money. It produces results faster than census studies.

2. Trained and experienced investigators conduct sample studies, which may improve measurement accuracy.
3. Only sampling works for infinite populations.
4. Sampling helps in estimating sampling errors and helps to learn about population characteristics.

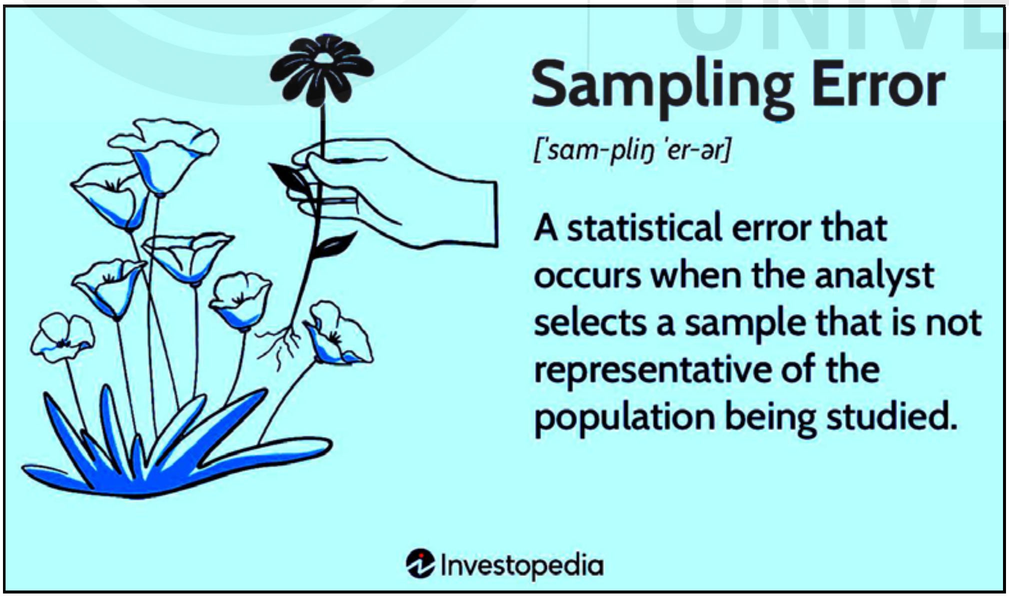
After understanding the significance of sampling, it's time to understand the concept of population and sample in more detail. However, the basic concept was discussed earlier also. The detailed discussion on population and sample is given in the subsequent section.

8.3.3 Sampling & Non-Sampling Errors

When doing research, it's important to know all about errors that could happen during the data collection and review phase to make sure that the results are correct and reliable. There are two main types of research errors: sampling errors and non-sampling errors.

When doing research, it's important to think about both sampling and non-sampling errors. Each has its own effects on the accuracy and dependability of the results. Sampling errors happen because you are only looking at a group and not the whole population. You can reduce these errors by using the right sampling methods. Other types of errors, called non-sampling errors, are caused by a bigger range of issues in the research process and need strict quality control measures to be fixed.

While both types of errors can affect how accurate the results are, they come from different sources and need different ways to be fixed. For reliable and trustworthy study results, it is important to fully understand the differences between these two types of mistakes and how to use strategies to lessen their effects.



Source: Investopedia **Figure 8.3 Sampling error**

The data collected by sample surveys represents only a fraction of the entire population, so it is inherent that some of it may lack accuracy. This

inaccuracy is generally referred to as sampling error or error variance. Sample errors refer to inaccuracies that incur as a result of the sample procedure. Given random sampling, these mistakes refer to random fluctuations in the sample numbers that closely approximate the actual population values.

Sampling errors can occur when a study is conducted on a subset of the population rather than the entire population. These errors occur due to the selection of a limited segment of the population, resulting in a significant volume of variation. The presence of random sample selection invariably results in some degree of sampling error, regardless of the proper execution of the sampling plan. For instance, when you select a sample of 500 individuals from a population of 10,000, it is possible that this sample may not precisely mirror the attributes of the entire population. This may lead to discrepancies between the sample outcomes and the actual characteristics of the population. By employing more sophisticated sampling techniques such as stratified or cluster sampling, or by increasing the sample size, the sampling error can be quantified and minimised.

On the other hand, non-sampling errors have nothing to do with sampling. They arise from other stages of the research process. This type of error can occur at any stage, from data collection to data processing, and it may be more difficult to detect and correct. Some of the issues that are not caused by sampling include incorrect measurements, poor data entry, nonresponse bias, and interviewer bias. *For example, if those who respond to a poll question do not comprehend it or if inaccurate data is captured, the ultimate findings will be erroneous and lead to incorrect conclusions*. Non-sampling errors can sometimes be more damaging than sampling errors since they might alter the overall outcome of a study by introducing bias.

The presence of random variation in the sample implies that it may not accurately represent the entire population. This is the underlying cause of sampling errors. To estimate the magnitude of these errors, it is common to express them as a confidence interval or margin of error. Conversely, non-sampling errors frequently occur unexpectedly and can be attributed to several factors such as poorly designed survey instruments, fatigued respondents, problems in data processing, or deliberate data falsification. Given the multitude of categories and degrees of complexity associated with non-sampling errors, it is imperative to engage in meticulous planning, meticulous data gathering, and comprehensive data verification in order to mitigate their impact.

Although the magnitude of sampling errors tends to decrease as sample sizes and sampling techniques improve, it is not always the case that non-sampling errors also decrease with an increase in sample size. Indeed, in certain instances, expanding the sample size might worsen non-sampling mistakes if the techniques used for data gathering or processing are defective. If, *for example, an online survey platform has a software defect that incorrectly interprets user inputs, augmenting the number of respondents will only result in a greater quantity of inaccurate data.*

When doing research, it is important to know that non-sampling mistakes need more proactive ways to be prevented and reduced than sampling errors. Sampling errors are often unavoidable and manageable. Researchers can cut down on sampling error by making sure their samples are as representative as possible and carefully planning how they will pick them. But lowering non-sampling mistakes needs careful attention to every part of the research process, from coming up with questions to making sure data collectors are trained and data accuracy is carefully checked.

Here are some important points about sampling errors:

i) Sampling errors are random variations in sample values that are compared to the real parameters of the population. These errors are compensatory because they happen at random and have an equal chance of going either way. This means that the expected value of these mistakes is zero.

ii) Increasing the size of the group lowers sampling error, which is not as important when the populations are similar.

iii) 'Precision of the sampling plan' is a standard way to measure sampling error for different sample sizes and patterns. More samples make the accuracy better. But there are some problems with making the group bigger, like the fact that it costs more to collect data and makes systematic bias more likely. So, the best way to improve accuracy is to pick a sampling plan that has a lower sampling error for the sample size and cost incurred. In fact, though, people like less precise designs because they are easier to use and make it easier to deal with systematic bias. When academics choose a sampling method, they need to make sure that it produces a sampling error that isn't too big and helps them deal with systematic bias better.

iv) Sampling errors happen by chance and can go either way. The nature of the universe determines how big the sampling error is. The smaller the sampling error, the more uniform the universe is.

v) Sampling error goes down as sample numbers get bigger, and the opposite is also true. It is possible to find a way to measure random sampling error for a certain sample design and number. This is called sampling plan precision.

vi) To find the sampling error, multiply the critical value at a certain level of importance by the standard error.

In contrast to sampling errors, non-sampling errors can arise during the collection of actual data and are present in all surveys, whether they are census or sample surveys. There is currently no method available to quantify non-sampling mistakes. Non-sampling errors refer to inaccuracies in the research process that are not directly caused by the sample procedure. These errors can arise at any phase, encompassing data extraction, data manipulation, or data interpretation. Non-sampling mistakes are frequently more troublesome than sampling errors due to their potential to generate systemic biases that deviate from the study objectives.

The different types of non-sampling errors are briefly discussed below:

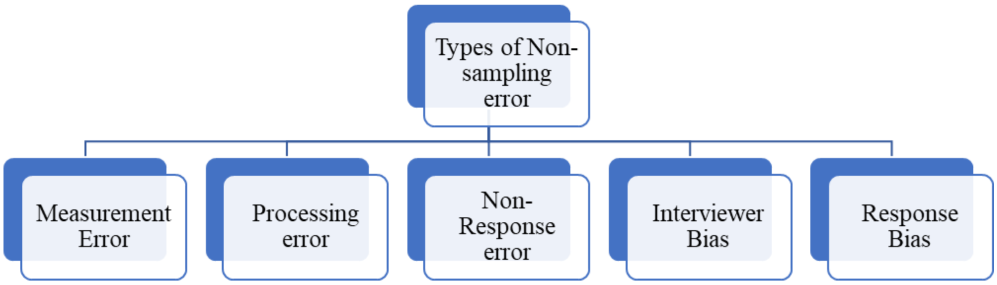


Figure 8.4: Types of Sampling Error

Measurement Errors: Occur when there is a flaw in the way data is collected or measured. This can be due to poorly designed survey questions, faulty equipment, or respondent misunderstanding.

Processing Errors: Arise during data processing, such as coding errors, data entry mistakes, or errors in data analysis.

Nonresponse Errors: Occur when a significant portion of the selected sample does not respond, leading to a bias if the non-respondents differ systematically from respondents.

Interviewer Bias: Introduced when the behaviour or characteristics of the interviewer influence the responses of participants.

Response Bias: Occurs when respondents provide inaccurate answers, either intentionally or unintentionally, due to social desirability, recall issues, or question wording.

Non-sampling errors have a substantial impact on research since they might skew the data and lead to incorrect findings. Although sampling errors often diminish with larger samples, non-sampling errors can persist regardless of sample size and may grow if not treated effectively. To reduce non-sampling errors, researchers must carefully design data collection devices, provide extensive training to data collectors, pre-test survey tools, and develop stringent data validation procedures. To reduce these errors, it is also critical to use consistent and dependable data collection and analysis methodologies.

8.3.4 Sampling Frame

A sampling frame is a fundamental concept in research methodology that forms the basis for choosing a sample from a specified population. It is essentially a comprehensive inventory or database that contains all the components or entities within the population from which the sample will be selected. Strict precision and comprehensiveness of the sampling frame are essential as they immediately impact the representativeness of the sample and, thus, the credibility of the study results.

Sample frame refers to the fundamental unit or group of units that might serve as the foundation for a sample operation, commonly known as sampling units. Therefore, a sample frame can be described as an exhaustive compilation encompassing all possible sampling units. Hence, it can be

asserted that a sampling frame is a collection of components used to select the sample. If the population is limited and the time frame is either the current or historical period, then it is achievable for the frame to be completely the same as population. In most cases, they cannot be distinguished because it is not possible to directly sample from the population. Therefore, this frame is either produced by a researcher expressly for their study or may incorporate an existing roster of the community. One appropriate framework for carrying out an opinion survey in a city is the telephone directory. Irrespective of the chosen frame, it should precisely represent the population.

The sampling frame and sampling design are connected in a way that makes the sampling design a unique plan for getting a sample from the sampling frame. A researcher's sample selection methodology is the exact plan or method they use to pick the units of sampling from which they draw conclusions about the whole community. Choices are made about the sampling strategy before the data are collected.

So, a Sampling Frame is a full list of all the things, people, or units that make up the community being studied. It's the operational population, which is where a sample is picked for study purposes. A sampling frame is very important because it gives you a clear and organised way to pick a sample that is a good representation of the whole community. When the sampling frame is carefully thought out, it makes sure that every part of the population has an equal chance of being chosen for the sample.

Following are the Characteristics of a Good Sampling Frame:

1. **Completeness:** The sample frame must encompass all individuals in the population without any exclusions or repetitions. Incomplete frames might result in imbalanced samples that fail to properly represent the population.

2. **Latest Information**: The information contained within the sample frame must be latest. The use of obsolete data may lead to the inclusion of non-existent or irrelevant units, therefore undermining the integrity of the sample.

3. **Accuracy:** The information included in the sample frame must be exact and accurately represent the attributes of the population. Flaws can result in inaccuracies in the selection of samples and, eventually, in the conclusions of the study.

4. **Appropriate Scope:** The sample frame should align with the focus of the investigation. The inclusion of units should be limited to those that are pertinent to the research, while excluding those that are not.

Some Examples of Sampling Frames are as follows:

- **Individual Lists:** These can be voter registration lists, customer files, job rosters, or hospital patient records, among other things.

- **Geographical Areas**: Maps or lists of households in a certain area can be used to set up a sampling frame for study that is based on an area.

It is not always possible to generate the best sampling frame, there are some challenges with the development of sampling frames, some of them are as follows:

1. **Under coverage:** The phenomenon of under-coverage arises when certain individuals within the population are not encompassed within the sampling frame, resulting in their exclusion from the sample. Over-coverage occurs when the sampling frame includes units that are not present in the target population, therefore potentially adding extraneous data into the sample.

2. **Outdated Information:** Insufficient updates to the sample frame may result in the inclusion of obsolete information, such as contact details, which can harm the efficiency of data gathering and introduce possible biases.

3. **Duplication:** Duplication within a sample frame refers to the redundant listing of certain units, which raises the likelihood of their selection and introduces bias.

In order to improve the Sampling Frames, one needs to put focus on the factors listed below:

1. **Periodic Updates:** Assure that the sample frame is consistently revised to accurately represent the present condition of the population.

2. **Verification and Cleaning:** Regularly validate and cleanse the sample frame to eliminate duplicate entries and rectify any inconsistencies.

3. **Multiple Sources:** Synthesise data from several sources to establish a more thorough and precise sample frame.

In this section, we learnt that a sampling frame is an important part of research because it helps choose a group that is representative of the whole population. The sampling frame's quality determines how accurate and reliable the sample is, which in turn determines how valid the study findings are. For researchers to be sure that their studies are successful, they must carefully build and keep up a high-quality sampling frame.

8.4 METHODS OF SAMPLING

Sampling is an essential component of research methodology, especially when working with very large populations, where it is not feasible to analyze every individual. The fundamental practice of sampling is the deliberate selection of a smaller group from a larger population in order to make conclusions about the whole population. The procedure of sampling involves a set of predetermined processes, which we have discussed earlier in this unit in section 8.4.1.

Once you understand how sampling works, you need to understand about the different types of sample methods. There are different types of sample methods that are based on two main factors: the representation basis and the element selection basis.

1. On element selection basis: depending on how the elements are chosen, the sample can be either open or closed. This type of sampling is called "unrestricted sampling," and it means that each part of the sample is chosen at random from the whole community. All other types of sampling are called "restricted sampling."
2. On the representation basis: if you look at the sample based on representation, it could be probability sampling or non-probability sampling.

The selection between the two groups is crucial since it directly impacts the selected sample strategy. A meticulously devised sampling strategy ensures that the sample accurately reflects the entire population. By reducing the likelihood of bias, the accuracy of the research findings is enhanced. Furthermore, it enhances the efficiency of the study process, results in time and cost savings, and provides precise data for analysis.

Majorly the types of sampling are broadly classified into two categories:

1. Probability Sampling
2. Non-Probability Sampling

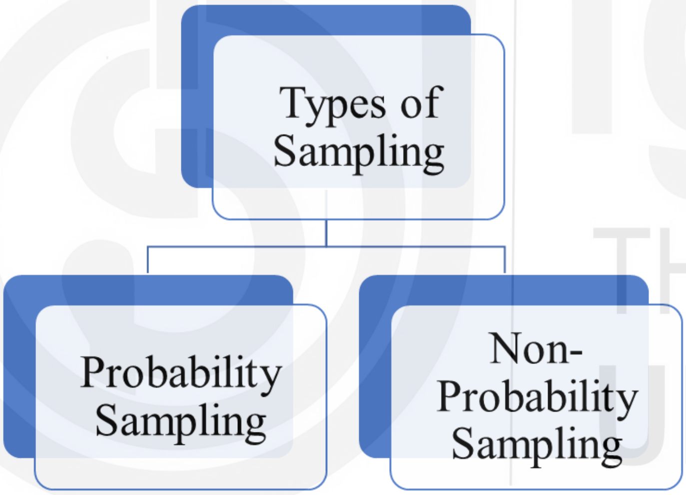


Figure 8.5: Types of Sampling

Random selection is the foundation of probability sampling, whereas non-random selection is the foundation of non-probability sampling. Sample methods can be broadly classified into two primary categories: non-probability sampling and probability sampling. The following section examines these two concepts individually.

8.4.1 Probability Sampling

Probability sampling, also known as random sampling or chance sampling, is a method of selecting samples. This sampling design ensures that every object in the universe has an equal probability of being selected for sampling. It is a lottery method where units are mechanically selected from the given group. The selection of an object is determined by blind chance. Probability or random sampling can be guaranteed based on probability, allowing us to precisely quantify estimating errors or significance, thereby making it more advantageous than selective sampling. Probability sampling ensures the

adherence to the law of Statistical Regularity, which states that a random sample would, on average, possess the same composition and characteristics as the larger population. For this reason, random sampling is superior for obtaining a representative sample. Random sampling from a finite population ensures that every possible combination of samples has an equal probability of being selected, and that every item in the population has an equal probability of being considered for sampling. Sampling without replacement refers to the selection of an item for the sample, which is then excluded from further selection. During sampling with replacement, the element is reintroduced into the population prior to the selection of the subsequent element. Prior to selecting the second element, it is possible for the same element to be present twice in the sample.

In brief, the implications of random sampling (or simple random sampling) are:

a) It ensures that every element in the population has an equal chance of being selected for the sample, and all choices are mutually independent.

b) It imparts an equal probability of selection to any conceivable sample combination.

In short, in probability sampling, each element within the population possesses a predetermined, non-zero probability of being chosen. The preference for this method frequently stems from its capacity to generate impartial samples, therefore enhancing the generalisability of the findings.

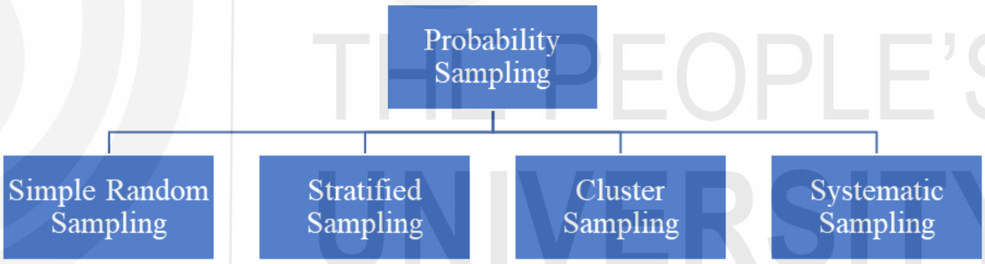


Figure 8.6 Types of Probability Sampling

Some prominent methods of probability sampling include:

1. **Simple Random Sampling:** In this every individual in the population has an equal chance of being selected. The selection is purely random, often achieved through lottery methods or random number generators.

   How it Works: A complete list of the population (sampling frame) is required, and each member is assigned a unique number. Numbers are randomly drawn to select participants.

   Example:

   a. *A company wants to survey employee satisfaction. If the company has 1,000 employees, they list all names and use a random number generator to select 100 employees for the survey.*

   b. *In a medical study, researchers randomly select 50 participants from a list of 500 patients.*

Advantages: Reduces bias, ensures fairness.

Challenges: Requires a complete and accurate sampling frame, which may not always be available.

2. **Stratified Sampling:** In this the population is divided into subgroups or strata based on shared characteristics, such as age, income, or education level. A random sample is then taken from each subgroup.

   How it Works: Identify strata, ensure they are mutually exclusive, and then use random sampling within each stratum.

   Example:

   a. *In a national election survey, the population is divided into age groups: 18-30, 31-50, and 51+. A random sample of voters is selected from each age group to ensure that all age categories are represented.*

   b. *A company studying product preferences divides customers by income brackets (low, middle, high) and selects a sample from each bracket.*

   Advantages: Ensures representation of key subgroups, improves the precision of estimates.

   Challenges: Requires detailed knowledge of population characteristics to form strata accurately.

3. **Cluster Sampling:** In this Sampling technique the population is divided into clusters, often based on geographical or organizational criteria. Entire clusters are randomly selected, and all members within these clusters are surveyed.

   How it Works: Identify clusters, ensure clusters are representative of the population, and then randomly select clusters for study.

   Example:

   a. *A researcher studying school performance randomly selects 10 schools out of 100 in a district and surveys all students in the selected schools.*

   b. *A marketing firm conducting a city-wide survey divides the city into neighbourhoods (clusters) and randomly selects 5 neighbourhoods to collect data from all households.*

   Advantages: Cost-effective, suitable for large populations spread across wide areas.

   Challenges: Risk of higher sampling error if clusters are not representative

4. **Systematic Sampling:** In this sampling technique a random starting point is selected, and every nth member of the population is chosen. This method works best with an ordered population list.

   How it Works: Determine the sampling interval (k = population size ÷ sample size). Start at a random point and select every kth individual.

Example:

In a factory quality check, inspectors randomly pick the 3rd item on a conveyor belt and then select every 10th item for inspection.

For a study on college students' dietary habits, researchers list students alphabetically and select every 5th student.

Advantages: Simple, time-efficient, ensures systematic coverage.

Challenges: If the population has a periodic pattern (e.g., conveyor belt items are arranged in repeating cycles), it may introduce bias

8.4.2 Non-Probability Sampling

A non-probability sampling method is unable to determine the probability of sampling each item in the source population. Selective, purposive, and judgemental sampling are non-probability sampling techniques. In purposive sampling, the researcher deliberately selects the sample items, and this selection is considered definitive. Non-probability sampling is the deliberate selection of particular units of the universe for sampling, based on the notion that a small mass from a larger one will accurately represent the whole. In order to examine the economic situation in a state, it is possible to select a small number of towns and villages for thorough analysis that serve as a representative sample of the state. Thus, this sample design depends on the viewpoint of research organisers. Under such a design, sample selection may be influenced by personal considerations. The investigator may selectively choose a sample that serves to bolster his stance, therefore contaminating the whole study. Every sampling technique carries the inherent possibility of bias provided that the investigators maintain neutrality, operate without prejudice, and possess the necessary expertise to make sound judgements, the outcomes of a meticulously chosen sample analysis might be precise. Nevertheless, such a sample cannot ensure the inclusion of every element. This sampling technique is incapable of evaluating sample error, so bias, whether significant or minor, is always inherent. This sampling methodology is seldom employed in significant research trials. Given its efficiency in terms of time and money, this sampling method can be employed for small enquiries and research conducted by individuals. Non-probability sampling encompasses the use of quotas. Quota sampling involves assigning interviewers strata-based quotas to fill with different specified limits. Interviewers select survey questions at their own discretion. This sample technique is both convenient and cost-effective. The chosen samples lack randomness. The inferences of quota samples cannot be statistically analysed as they are judgement samples.

A short explanation of non-probability sampling is that samples are chosen based on personal opinion instead of chance. These methods are easy to use, but they are not as good at applying results to the whole population.

Some prominent methods of non-probability sampling include:

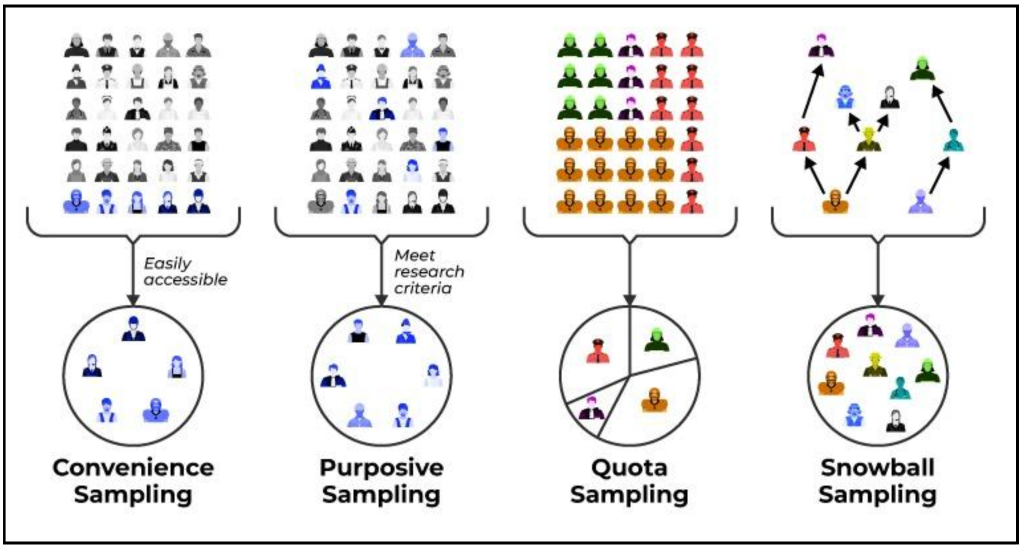


Figure 8.7: Types of Non Probability Sampling

Convenience Sampling: Under this sampling method samples are selected based on ease of access and proximity to the researcher. It is commonly used in exploratory research or when there are constraints in time, resources, or availability of participants.

How it Works: Participants are chosen simply because they are convenient to recruit or readily available, without regard to randomness or representativeness.

Example:

a. *A researcher conducting a study on students' study habits surveys students sitting in the campus library because they are easily accessible.*

b. *A small business surveys customers visiting a particular store during a weekend to understand their satisfaction level.*

Advantages: Quick, inexpensive, and easy to conduct.

Challenges: Results may not be representative of the broader population, introducing selection bias.

Purposive or judgemental sampling: In this method, the researcher selects participants based on their knowledge, judgment, and the purpose of the study. This method focuses on those individuals who are believed to provide the most relevant or useful data.

How it Works: The researcher identifies criteria or characteristics that are important for the study and chooses participants who meet those criteria.

Example:

a. *In a study on the effectiveness of teaching methods, a researcher selects only teachers with more than 10 years of experience in innovative pedagogy.*

b. *For a qualitative study on the experiences of cancer survivors, the researcher purposefully chooses individuals who have undergone treatment in the past 5 years.*

Advantages: Ensures that participants are relevant to the research objectives, often used in qualitative research for in-depth insights.

Challenges: Prone to researcher bias, as the selection is subjective and may exclude diverse perspectives.

Quota Sampling: In this method population is divided into specific categories or quotas and ensures that the sample includes a certain number of participants from each category. However, within these quotas, participants are not selected randomly.

- How it Works: The researcher identifies key characteristics (e.g., age, gender, income) and sets quotas to reflect these proportions in the sample.
- ***Example:***
  a. *In a study on consumer behaviour, researchers aim to include 50% males and 50% females in the sample. They continue recruiting participants until these quotas are met, regardless of randomness.*
  b. *A political survey ensures that 30% of respondents are aged 18–30, 40% are 31–50, and 30% are above 50, matching the population's age distribution.*
- **Advantages:** Ensures representation of key population segments, quicker and less expensive than probability sampling methods.
- **Challenges:** Selection bias may occur as participants within each quota are not chosen randomly.

Snowball sampling: In this method, existing participants recruit future participants from among their acquaintances. It is often used for studying hard-to-reach or hidden populations.

How it Works: The researcher identifies initial participants (seeds), who then refer others to join the study. This process continues until the desired sample size is reached.

Example:

a. *A study on the experiences of undocumented immigrants starts with a few initial participants who then connect the researcher to others in their community.*

b. *Researchers studying drug addiction begin with known addicts and rely on them to introduce others in similar circumstances.*

Advantages: Effective for accessing hard-to-reach population or groups that are not easily identified through traditional sampling methods.

Challenges: Potential for bias if participants refer to individuals who are similar to themselves, leading to limited diversity in the sample.

Based on the analysis conducted in this section, it can be established that a Sampling design is a fundamental component of research technique. Through

meticulous selection and implementation of the suitable sampling technique, researchers can provide dependable and generalisable findings that augment the progress of knowledge in their respective sector. Gaining a comprehensive understanding of the advantages and constraints of various sampling methods is essential for making well-informed judgements during the research process. Moreover, it is crucial to employ consistent and dependable techniques for gathering and analysing data in order to minimise these errors.

Check Your Progress A:

1) What is the difference between Statistics and Parameter?

…………………………………………………………………………………
…………………………………………………………………………………
…………………………………………………………………………………
…………………………………………………………………………………
…………………………………………………………………………………

2) Explain the Significance of Sampling.

…………………………………………………………………………………
…………………………………………………………………………………
…………………………………………………………………………………
…………………………………………………………………………………
…………………………………………………………………………………

3) What is a Sampling Frame? Explain the characteristics of a good sampling frame.

…………………………………………………………………………………
…………………………………………………………………………………
…………………………………………………………………………………
…………………………………………………………………………………
…………………………………………………………………………………

4) What is sampling error? How can it be minimised?

…………………………………………………………………………………
…………………………………………………………………………………
…………………………………………………………………………………
…………………………………………………………………………………
…………………………………………………………………………………

8.5 SAMPLING UNIT

The sample unit is a crucial notion in research that describes the specific entities from which data is collected. Its appropriate identification and definition are vital for ensuring that the sample is representative of the

population and that the study findings are accurate and reliable. By understanding and precisely selecting the sampling unit, researchers can boost the quality of their studies and provide relevant insights into their research issues.

In research, the concept of a sampling unit is fundamental to the process of acquiring and interpreting data. An individual object or element from which data is gathered and observations are produced is referred to as the sampling unit. It serves as the foundation for selecting a sample from the larger population and plays a significant role in evaluating the accuracy and representativeness of the research findings.

It can be specified that a sampling unit is the basic unit of analysis within the sample frame. It might vary greatly depending on the type of the research and the topic of the investigation. *For example, in a study on individual health behaviours, each person may be a sampling unit*. Conversely, in research examining organizational practices, entire organization might function as the sampling unit. The value of the sampling unit resides in its impact on the research design and data gathering method. Accurately identifying and defining the sampling unit guarantees that the sample drawn is representative of the population and that the data obtained can be effectively evaluated.

Types of Sampling Units

1. Individual Units: These individual units are the fundamental type of sampling units that reflect individual entities within the population. *Example include individuals, animals, or standalone objects such as products or pieces of equipment.*

2. Group units are employed when the research emphasis is on groups rather than individuals. Example include residences, educational institutions, or enterprises.

3. Geographical units are utilised in research that pertains to geographical locations or regions. *Example include Urban centres, residential areas, or precisely defined geographical regions.*

A very important part of the study process is carefully choosing the sampling units. The choice should be based on the goals of the study and the characteristics of the group being studied. Selecting the right sampling groups is important to make sure that the sample really does reflect the whole population and that the data collected is useful and relevant.

It is important to stress that choosing a sampling unit comes before choosing a sample. The units used for sampling can be people, families, clubs, schools, or physical units like a state, district, or hamlet. They can also be building units like a house or flat. The researcher has to pick at least one of these groups in order to do his study.

Based on the analysis in this section, it is clear that the sampling unit should be chosen in a way that makes sure the sample is a good

representation of the whole community. This helps make sure that the results are correct and reliable. It also changes the methods and logistics of collecting data depending on the sampling unit chosen. In particular, getting information from people might require different methods than getting information from businesses.

8.6 SAMPLING DISTRIBUTION

In statistics, a sampling distribution is the probability distribution of a specific statistic produced from numerous random samples drawn from the same population. The expected value distribution is the set of values that a statistic, such as the mean or proportion, would take if a large number of samples of the same size were drawn from the population. This concept allows researchers to understand the degree of fluctuation and dependability of sample statistics, as well as assess their correctness in estimating population parameters.

Figure 8.8: Sampling Distribution

A Sampling distribution is an important topic in statistical study because it allows us to draw inferences about a population based on data from a sample. A sampling distribution is a statistical distribution that depicts the distribution of a certain statistic, such as the sample mean or sample proportion, derived by periodically sampling a population. The framework provides a way for understanding the differences in sample statistics between different samples, which is required for hypothesis testing and population parameter calculations.

Examples

1. **Sample Mean Distribution:** The sample mean distribution refers to the distribution of sample means obtained by routinely selecting samples of a certain size from a population and computing the mean for each sample. Under the Central Limit Theorem, this distribution will tend to a normal distribution when the sample size is sufficiently large.

2. **Sample Proportion Distribution:** The sample proportion distribution is a statistical measure used to estimate the population proportion and

evaluate the variability of this estimate when analysing proportions, such as the percentage of a population that supports a certain policy.

It is important to understand the sampling distribution as it contributes to the following:

1. **Estimation:** The estimation of population parameters relies heavily on the selection of sampling distributions. Through the analysis of the sampling distribution, researchers are able to compute confidence intervals in order to approximate the probable location of the true population parameter.

2. **Hypothesis Testing:** The sampling distribution is used in hypothesis testing to ascertain if the observed sample statistics deviate significantly from the expected values of a null hypothesis. It serves as the foundation for computing p-values and formulating conclusions on hypotheses.

3. **Understanding Variability:** The notion of sampling distribution enables researchers to measure the extent of variation in sample estimates. This phenomenon shows the fact that various samples taken from a common population can provide varied outcomes, offering valuable understanding of the accuracy and dependability of statistical estimations.

The key Features of the sampling distributions are listed below:

1. **Mean of the Sampling Distribution:** The mean of the sampling distribution of a sample statistic is equivalent to the estimated population parameter provided by the sample. As an illustration, the average of the sampling distribution of the sample mean is equivalent to the average of the population. Aforementioned characteristic guarantees that sample statistics serve as impartial estimators of population parameters.

2. **Standard Error:** The standard error is a statistical metric that quantifies the level of variability in the sample distribution. It quantifies the anticipated deviation of the sample statistic from the population parameter. The decreasing trend of the standard error with increasing sample size suggests that larger samples yield more accurate estimations of the population parameter.

3. **Shape of the Distribution:** The form of the sampling distribution is determined by both the sample size and the intrinsic population distribution. The Central Limit Theorem states that when the sample size is large, the distribution of the sample mean will closely resemble a normal distribution, irrespective of the type of the demographic distribution. As the sample size increases, this approximation becomes considerably more precise.

Types of Sampling Distributions

1. **Sampling Distribution of the Sample Mean:** The most common type of sampling distribution is the sampling distribution of the sample mean. If we repeatedly take random samples of a fixed size from a population and calculate the mean for each sample, the distribution of these means will form the sampling distribution of the sample mean. According to the

Central Limit Theorem, if the sample size is sufficiently large, this distribution will be approximately normal, regardless of the population's distribution.

2. **Sampling Distribution of the Sample Proportion:** When dealing with categorical data, we often look at the proportion of observations that fall into a particular category. The sampling distribution of the sample proportion is the distribution of the proportions of a given category across different samples. Like the sample mean, if the sample size is large enough, the distribution of the sample proportion will also be approximately normal.

3. **Sampling Distribution of the Sample Variance:** The sampling distribution of the sample variance refers to the distribution of variances computed from different samples drawn from the same population. Unlike the mean and proportion, the distribution of the sample variance is typically skewed, especially for small sample sizes. As the sample size increases, the distribution becomes more symmetric.

4. **Sampling Distribution of the Sample Standard Deviation:** Similar to the sampling distribution of the sample variance, the sampling distribution of the sample standard deviation involves the spread or dispersion of data within a sample. The distribution of the sample standard deviation is also skewed for small samples but becomes more normal as the sample size increases.

5. **t-Distribution:** When dealing with small sample sizes (typically less than 30), and the population standard deviation is unknown, the sampling distribution of the sample mean follows a t-distribution rather than a normal distribution. The t-distribution is similar to the normal distribution but has heavier tails, meaning there is a higher probability of extreme values.

6. **Chi-Square Distribution:** The chi-square distribution is relevant when considering the sampling distribution of the sample variance. It is used in hypothesis testing and in constructing confidence intervals for population variances. The chi-square distribution is skewed to the right, and its shape depends on the degrees of freedom associated with the sample.

7. **F-Distribution:** The F-distribution arises when comparing the variances of two samples. It is used in analysis of variance (ANOVA) and in tests that compare the variability of two or more groups. The F-distribution is also skewed to the right and is characterized by two sets of degrees of freedom.

It is important to understand these kinds of sampling distributions for statistical inference, which lets us use sample data to determine population parameters, test hypotheses, and make predictions.

The discussion in this part leads to the conclusion that the idea of sampling distribution is important for knowing how sample statistics relate to population parameters and for drawing correct statistical conclusions. It can be used to determine population parameters, test hypotheses, and figure out

how different sample estimates are from each other. Researchers can be sure that their findings are based on strong statistical principles and correctly reflect the characteristics of the population by looking into sampling distributions.

8.7 CENTRAL LIMIT THEOREM

The Central Limit Theorem says that if you pick large number of random samples from a population with a fixed mean and variance, the distribution of the sample means will look a lot like a normal distribution. The accuracy of this estimate gets a lot better as the sample size increases. The Central Limit Theorem (CLT) is one of the most important ideas in statistics. It says that as the sample size increases, the distribution of sample means tend to converge to a normal distribution, regardless of how the original population was distributed. This theorem is one of the most important inferential statistics because it is the ground for many statistical methods and procedures.

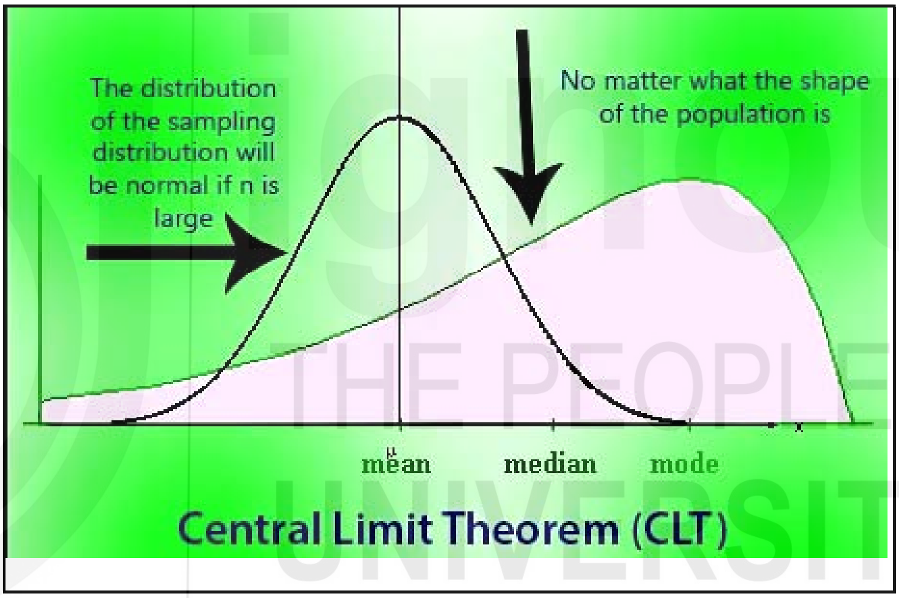


Figure 8.9: Central Limit Theorem

Key Components of Central Limit Theorem are:

1. **Sample Mean Distribution:**

   Mean of Sampling Distribution: The mean of the sampling distribution of the sample mean will be equal to the population mean (μ).

   Standard Error: The standard deviation of the sampling distribution (standard error) is given by $\sigma/\sqrt{n}$, where σ is the population standard deviation and n is the sample size.

2. **Shape of Distribution:**

   Normal Distribution: As the sample size increases, the sampling distribution of the sample mean will approach a normal distribution, even if the original population distribution is not normal. For most practical purposes, a sample size of 30 or more is considered sufficient for the CLT to apply.

The Central Limit Theorem (CLT) is a fundamental result in statistics that states that, given a sufficiently large sample size, the sampling distribution of the sample mean will be approximately normally distributed, regardless of the population's original distribution. The mathematical expression for the Central Limit Theorem can be expressed as follows:

Let $X_1, X_2 \ldots \ldots X_n$ be a random sample of size n drawn from a population with mean μ and finite variance σ^2.

1. **Sample Mean:**

   The sample mean x̄ is given by:

   $$\bar{x} = \frac{1}{n}\sum_{i-1}^{n} X_i$$

2. **Standardized Form (Z-Score):**

   The Central Limit Theorem states that the standardized form of the sample mean approaches a standard normal distribution as the sample size n increases:

   $$Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \approx N \quad (0, 1)$$

Where N (0, 1) represents the standard normal distribution with a mean of 0 and a standard deviation of 1.

In simpler terms, as n increases or becomes large, the distribution of $\bar{x}$ will tend to be normally distributed with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$, even if the original population distribution is not normal. This result is extremely powerful because it allows for the application of normal probability theory to problems involving averages of independent random variables.

Some of the applications and importance of CLT are discussed below :

- **Estimation and Confidence Intervals:** The CLT lets you build confidence intervals around sample means. Researchers can use the properties of the normal distribution to predict population parameters and find confidence intervals because the sampling distribution of the mean is pretty close to normal.

- **Hypothesis Testing:** The CLT is used by many hypothesis tests, like the t-test and the z-test, to draw conclusions about population statistics. These tests can be used to see if the sample data are significantly different from what would be expected if the null hypothesis is true because the sampling distribution is normal.

- **Practical Use:** The CLT lets you use methods and techniques based on normal distribution in a number of real-life situations. It makes it easier to draw conclusions and make forecasts from sample data.

Illustration

Let us consider the task of calculating the mean height of adults residing in a city. Take several random samples of 50 adults each and compute the average height for each sample. As per the Central Limit Theorem (CLT), if the height distribution in the population is not normally distributed, the distribution of the sample means will converge to a normal distribution as the number of samples increases. Application of normal distribution-based approaches enables the estimation of the average height and the construction of confidence intervals.

The discussion presented in this section establishes that the Central Limit Theorem is a fundamental principle in statistical theory and application. It establishes a basis for several statistical methods by guaranteeing that the distribution of sample means approximately follows a normal distribution when the sample sizes are sufficiently large. This characteristic enables the calculation of population parameters, the testing of hypotheses, and the implementation of methods based on the normal distribution to real-world data.

8.8 DETERMINATION OF SAMPLE SIZE

Determining an optimal sample size is essential to guarantee the reliability, validity, and generalisability of the findings of any study. The magnitude of the sample size directly impacts the accuracy of the estimations and the statistical power of any statistical tests. The determination of sample size for a research project involves consideration of various elements and methodologies. Some of them are enumerated below:

1. **Purpose of the Study:**

   Estimation: For estimating population parameters (e.g., mean or proportion) with a specified level of confidence and precision.

   Hypothesis Testing: For detecting a specified effect size with a desired power level.

2. **Population Variability:**

   Greater variability in the population requires a larger sample size to achieve accurate estimates. Variability can be measured by the population standard deviation (σ) or variance (σ^2).

3. **Desired Confidence Level:**

   The confidence level indicates how confident you want to be that your estimate falls within a certain range of the true population parameter. Common confidence levels are 90%, 95%, and 99%. Higher confidence levels require larger sample sizes.

4. **Margin of Error (Precision):**

   The margin of error represents the range within which you expect the true population parameter to fall. A smaller margin of error requires a larger sample size.

5. **Effect Size:**

   In hypothesis testing, the effect size is the magnitude of the difference you want to detect. Smaller effect sizes require larger sample sizes to detect with adequate power.

6. **Power of the Study:**

   Power is the probability of correctly rejecting the null hypothesis when it is false (i.e., detecting an effect if one exists). Higher power requires larger sample sizes.

Sample size calculation is essential in statistics to determine the number of observations or replicates needed to achieve a desired level of precision or confidence in the results. The calculation depends on the desired confidence level margin of error, population variability, and, for certain studies, the effect size. The **Basic formula for Sample Size Calculation is discussed below:**

For a simple random sample when estimating a population mean, the sample size n can be calculated using the following formula:

$$n = \left(\frac{Z_a/2 \cdot \sigma}{E}\right)^2$$

Where

- $Z_{a/2}$ is the critical value from the standard normal distribution corresponding to the desired confidence level.
- σ is the population standard deviation (if unknown, an estimate can be used).
- E is the margin of error (the maximum allowable difference between the sample mean and the population mean).

Example

Let's calculate the sample size needed to estimate the average height of adult males in a city, with 95% confidence level and a margin of error of 2 cm. Assume the population standard deviation σi 8 cm.

1. Determine the confidence level:
   - For a 95% confidence level, $Z_{a/2} = 1.96$.
2. Specify the margin of error (E):
   - $E = 2$ cm.
3. Specify the population standard deviation (σ)
   - $\sigma = 8$ cm.
4. Plug the values into the formula:

$$n = \left(\frac{1.96 \cdot 8}{2}\right)^2$$

5. Calculate the sample size:

$$n = \left(\frac{15.68}{2}\right)^2 = (7.84)^2 = 61.47$$

Since the sample size n must be a whole number, you would round up to the next whole number:

$$n = 62$$

From the obtained result it can be interpreted that you would need a sample size of 62 adult males to estimate the average height with a margin of error of 2 cm at a 95% confidence level.

This section concludes with the learning that Determining the appropriate sample size involves the consideration towards the purpose of study, population variability, desired confidence level, margin of error, effect size, and power. Using statistical formulas and tools, researchers can calculate the required sample size to ensure accurate and reliable results. Practical constraints and pilot studies can further refine sample size estimates to balance statistical rigor with feasibility.

Check Your Progress B:

1) What is Sampling Distribution? Explain its key features.

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

2) What is the Central Limit Theorem?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

3) What are the applications of the Central Limit Theorem?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

4) What are the factors to be considered for determining sample size?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

5) State whether which of the following statement is True and False

   i) A sampling frame is a comprehensive list of all elements in a population from which a sample is drawn.

   ii) Non-probability sampling ensures that every member of the population has an equal chance of being selected.

   iii) Sampling errors can be reduced by increasing the sample size.

   iv) The Central Limit Theorem applies only when the population is normally distributed.

   v) Non-sampling errors can occur during data processing and data collection stages.

6) Mark the correct answer

   i) Which of the following is NOT a characteristic of a good sample?

      A. Representativeness

      B. Independence

      C. Convenience

      D. Adequate size

   ii) In probability sampling:

      A. Samples are selected based on convenience.

      B. Every member has a non-zero chance of selection.

      C. Bias is intentionally introduced.

      D. Only small populations are studied

   iii) The Central Limit Theorem states that the sampling distribution of the sample mean:

      A. Is always skewed

      B. Approaches a normal distribution as sample size increases

      C. Is independent of sample size

      D. Does not apply to large populations

   iv) The margin of error is:

      A. The range within which the true population parameter is expected to fall

      B. The probability of selecting a biased sample

      C. The variability in the population

      D. The critical value in hypothesis testing

8.9 LET US SUM UP

This unit provides an in-depth exploration of sampling as a fundamental aspect of research methodology. It begins with an overview of the importance of sampling, emphasizing its ability to enable researchers to study large populations efficiently and accurately without the need to examine every individual. Sampling is highlighted as a cost-effective and time-saving

approach that improves data quality and supports inferential statistics, enabling generalization of results to population.

The unit defines essential concepts, such as "population" (the entire group under study) and "sample" (a representative subset), along with their interrelationship, which is critical for ensuring research validity. It delves into the significance of sampling frames, stressing that an ideal sampling frame should be comprehensive, current, accurate, and aligned with the research's focus.

Key principles, such as the characteristics of a good sample representativeness, adequate size, randomness, and independence are discussed to ensure the reliability and applicability of research findings. The distinction between probability and non-probability sampling methods is also addressed. Probability sampling methods, like simple random sampling and stratified sampling, are shown to reduce bias and enhance representativeness, while non-probability methods, such as convenience sampling, are acknowledged for their practicality in specific contexts despite potential biases.

The unit elaborates on sampling and non-sampling errors, differentiating between errors arising from sample selection and those stemming from other stages of research, like data collection or processing. Strategies to minimize these errors are suggested, including careful design and rigorous quality control.

The concept of sampling distribution is introduced, explaining its role in understanding variability and reliability in sample statistics, with reference to the Central Limit Theorem (CLT). The CLT is underscored as a cornerstone of statistical inference, facilitating the application of normal distribution methods to sample data.

Lastly, the unit provides guidelines for determining sample size, considering factors like population variability, desired confidence levels, and margin of error. It emphasizes the balance between statistical rigor and practical constraints to ensure reliable and valid research outcomes.

Overall, this unit equips researchers with comprehensive knowledge to design robust sampling strategies, mitigate errors, and derive meaningful insights from their studies.

8.10 KEYWORDS

Central Limit Theorem (CLT): A statistical theory stating that the sampling distribution of the sample mean approaches a normal distribution as the sample size increases, regardless of the population's distribution.

Confidence Level: The probability that the true population parameter lies within the margin of error, commonly expressed as percentages like 90%, 95%, or 99%.

Non-Probability Sampling: A sampling method in which not all population members have an equal chance of selection, often relying on convenience or judgment.

Non-Sampling Error: Errors not related to the sampling process, such as measurement errors, data entry mistakes, or respondent biases, that affect the accuracy of the research.

Population: The complete set of individuals, items, or events of interest in a research study, from which a sample is drawn**.**

Probability Sampling: A sampling method in which every member of the population has a known, non-zero chance of being selected, ensuring randomness and reducing bias.

Sample: A representative subset of a population used to infer information about the whole population.

Sampling: The process of selecting a subset (sample) from a larger population to gather information and draw conclusions about the entire population.

Sampling Error: The difference between the sample statistic and the true population parameter, arising from studying a subset instead of the entire population.

Sampling Frame: A list or database containing all elements of a population from which a sample is selected.

Sampling Unit: The individual elements or groups of elements considered for selection in a sample, such as individuals, households, or organizations.

8.11 ANSWER TO CHECK YOUR PROGRESS

Answer to Check Your Progress B

5 (i) True (ii) False (iii) True (iv) False (v) True

6 (i) C (ii) B (iii) B (iv) A

8.12 TERMINAL QUESTIONS

1) Discuss the significance of sampling in research. Also explain its advantages and limitations.
2) Describe the steps involved in developing a sampling design and explain the importance of each step.
3) Explain the difference between probability sampling and non-probability sampling with examples.
4) Explain the various types of errors in sampling. How can non-sampling errors be controlled effectively?
5) Define margin of error and its role in research.

6) Illustrate the importance of the Central Limit Theorem in statistical analysis with an example.

8.13 FURTHER READINGS

1. Lohr, S. L. (2021). *Sampling: design and analysis*. Chapman and Hall/CRC.
2. Malhotra, N. K. (2020). *Marketing research: an applied orientation*. pearson.
3. Kothari, C. R. (2004). Research methodology.
4. Green, R. H. (1979). *Sampling design and statistical methods for environmental biologists*. John Wiley & Sons.

Note	These questions are helpful to understand this unit. Do efforts for writing the answer of these questions but do not send your answer to university. It is only for your practice.

BLOCK 3
DATA ANALYSIS

BLOCK 3 DATA ANALYSIS

This is the third block of the course **"Business Research."** It focuses on data preparation, processing, presentation and equipping learners with the skills to analyse data effectively. This block introduces concepts like descriptive statistics, measures of central tendency and dispersion, and relationships among mean, median, and mode. Additionally, it covers hypothesis formulation and testing, parametric and non-parametric tests, as well as correlation and regression techniques, emphasizing their significance in research.

The block, titled **"Data Analysis,"** consists of three units, described in detail below:

Unit 9: Data Preparation and Presentation

This unit introduces learners to the importance of data preparation, processing, and presentation, tailored to the type of study being conducted. Learners will explore how to prepare and process data systematically and present it in a way that is both effective and easy to understand. The focus is on presenting data in formats that convey the intended information clearly to the audience.

Unit 10: Descriptive Statistics

This unit familiarizes learners with descriptive statistics, which are summary measures that provide insights into a dataset. It covers measures of central tendency (mean, median, mode) and measures of dispersion (range, variance, standard deviation). Learners will also understand the relationships among mean, median, and mode, enabling them to summarize and interpret data meaningfully.

Unit 11: Formulation and Testing of Hypotheses

This unit focuses on the formulation and testing of hypotheses, a crucial step in most research studies. It explains the concept of hypotheses, their significance, and the steps involved in hypothesis testing. Learners will also gain an understanding of parametric and non-parametric tests, their applications, and the distinction between them. Additionally, the unit covers correlation and regression techniques, along with their types, enabling learners to analyse relationships and make predictions based on data.

UNIT 9 DATA PREPARATION AND PRESENTATION

Structure

9.0 OBJECTIVES

After studying this unit, you will be able to:

- Understand the importance of preparing data for analysis;
- Comprehend how to clean and edit data to ensure accuracy;
- Explore methods to present data clearly through tables and visuals; and
- Develop skills to communicate research findings effectively.

9.1 INTRODUCTION

In today's digital age, data serves as the backbone of decision-making and strategic planning. However, raw data is like an uncut diamond that requires proper processing and presentation to reveal its true value. In the previous unit, we explored the principles of sampling design. Once the data is collected, it must undergo further processing to align with the research

objectives. This unit introduces the concepts of **data preparation** and **data presentation**, and also highlights their importance in transforming unstructured information into meaningful insights. Data preparation is a foundational step in research and analysis. It ensures that errors, inconsistencies, and missing values in raw data are addressed before beginning the analysis. This process improves the reliability and accuracy of results by providing a solid base for informed decision-making. *For instance, a marketing researcher analyzing the customer feedback must first clean the data to remove duplicate and incomplete entries to reach reliable conclusions..*

On the other hand, Data presentation is the art of communicating the findings effectively. Whether through tables, graphs, or charts, presenting data visually enhances its impact, making it easier for audiences to grasp complex patterns and trends. Imagine presenting quarterly sales figures to a business team—visuals like bar charts or line graphs can instantly highlight growth areas or challenges.

In this unit, we will discuss the significance of data preparation and the techniques involved, such as data cleaning and editing. You will also learn about various methods of presenting data, including tabular and visual formats.

9.2 MEANING OF DATA

Data refers to a collection of facts, figures, words, or observations that provide valuable information for analysis and decision-making. In the context of business research, data forms the basis for identifying patterns, solving problems, and making informed decisions. It acts as the cornerstone for deriving insights that improve processes, evaluate strategies, and drive success in various business domains.

Sources of Data

Data used in business research can originate from several key sources:

- **User Input**: Responses from surveys, feedback forms, or text-based entries provided by customers, employees, or other stakeholders.
- **Sensors and Devices**: Data collected from technological tools such as GPS systems, smart devices, or IoT (Internet of Things) sensors.
- **Business Operations**: Transactional data, inventory logs, or employee performance records generated during routine business activities.

Importance of Data in Business Research

Data plays a critical role in modern business research by providing a solid foundation for analysis and strategy formulation:

- **Identifying Trends**: Data helps businesses recognize patterns in customer behavior, market dynamics, and financial performance.
- **Supporting Decision-Making**: Quantitative and qualitative data guide organizations in making evidence-based decisions.

- **Enhancing Forecasting**: Reliable data enables accurate projections, helping businesses prepare for future challenges and opportunities.

Managing and Protecting Data

Effective data management ensures that data remains accessible, accurate, and secure for business research:

- **Data Organization**: Proper classification and storage ensure that data is easily retrievable and ready for analysis.
- **Privacy Protection**: Safeguarding data against unauthorized access and adhering to regulations like GDPR is crucial to maintain trust and compliance.

9.3 TYPES OF DATA

Now let's discuss the types of data, Following are the major types of data in business research:

- **Quantitative Data:** Quantitative data is all about numbers. Think of it as data that answers questions like "How much?" or "How many?" It's measurable and can be expressed in terms of quantity, which makes it ideal for comparisons and statistical analysis. *For example, if you're looking at sales data for the past year, the total revenue generated or the number of products sold falls under quantitative data.* In business research, this type of data is critical for identifying trends, forecasting future performance, or evaluating whether a strategy is working. Whenever you see charts or graphs with numbers on them, you're likely dealing with quantitative data.
- **Qualitative Data:** On the other hand, qualitative data focuses on descriptions and meanings rather than numbers. This type of data gives us the "why" and "how" behind certain phenomena. *For example, if you conduct interviews with customers about their experiences with a product, their feedback—such as "This product is very useful" or "The customer service was disappointing"—is qualitative data.* It helps us understand behaviors, preferences, and sentiments. In business research, qualitative data often forms the foundation for creating new ideas or improving existing services. It's the kind of data you'll encounter when analyzing survey comments, focus group discussions, or even social media reviews.
- **Structured Data:** Structured data is perhaps the most straightforward type of data to work with. It's neatly organized into rows and columns, making it easy to store and analyze. *Picture an Excel sheet with customer names, order numbers, and purchase amounts—that's structured data.* In business research, this data is used for tasks like tracking performance metrics, generating reports, or performing in-depth statistical analysis. Because it's so well-organized, structured data is the backbone of most databases and is often the starting point for research projects.

- **Unstructured Data:** Now, let's move to the opposite end of the spectrum: unstructured data. This type of data doesn't have a defined format, which can make it challenging to analyze. However, it often holds rich insights that can be incredibly valuable for businesses. Examples include social media posts, emails, or videos. Imagine analyzing tweets about a brand during a product launch—each tweet may have different wording, tone, or length, but together, they can reveal public sentiment. Businesses use unstructured data to track brand reputation, understand customer preferences, and even detect potential risks. While it's harder to work with, advancements in technology, such as AI, have made unstructured data analysis more accessible.

- **Semi-Structured Data:** Semi-structured data is the middle ground between structured and unstructured data. It has some organization but doesn't fit perfectly into a table. Think about emails: they have structured headers (like the sender, recipient, and time sent) and unstructured content (the actual message). In business research, semi-structured data often acts as a bridge, helping researchers combine the clarity of structured data with the insights from unstructured data. For instance, analyzing website logs can reveal user behavior patterns, even if the data isn't fully structured.

- **Metadata:** Have you ever noticed the details about a file, like its creation date or file size? That's metadata—data about data. It doesn't give you the main content but instead provides context or additional details. For example, if you're analyzing a dataset of customer complaints, the metadata might include the date each complaint was logged and the type of complaint. In business research, metadata is incredibly useful for organizing large volumes of information. It helps researchers quickly locate and classify data, making it easier to focus on the actual analysis.

- **Big Data:** Finally, let's talk about big data. This is a term you've probably heard a lot, and for good reason—it's transforming how businesses operate. Big data refers to massive datasets that are too large and complex to be handled by traditional methods. These datasets often come from real-time sources like social media, IoT devices, or online transactions. For example, an e-commerce company might analyze millions of customer interactions to personalize product recommendations. In business research, big data allows us to identify patterns, predict trends, and make data-driven decisions on a scale never seen before. It's the foundation of technologies like AI and machine learning, which are reshaping industries.

9.4 SIGNIFICANCE OF DATA PREPARATION

Data preparation is a crucial step in any research or analysis process. It involves organizing, cleaning, and structuring raw data to make it suitable for analysis. Without proper preparation, data can contain errors, inconsistencies, or missing values that may lead to inaccurate conclusions or flawed results.

For example, if a company analyzes sales data with errors or missing values, it might make flawed business decisions, leading to financial losses.

The significance of data preparation lies in ensuring the reliability and validity of your analysis. Well-prepared data helps researchers save time, reduces errors, and enhances the quality of insights. It also makes complex datasets easier to manage and interpret. This step is not just about making data "usable" but also about building a solid foundation for effective decision-making and presentation.

In addition, data preparation allows researchers to identify patterns and relationships in the data more effectively. When raw data is cleaned and organized, it becomes easier to spot trends, outliers, or areas that require further investigation. For example, structured sales data can help businesses identify peak shopping periods, popular products, or regions with low sales, enabling them to create targeted strategies. By investing time in this step, you can ensure that the subsequent analysis is not only accurate but also actionable, leading to well-informed decisions and better outcomes.

9.5 PROCESSING OF DATA

Processing of data is a key step that transforms raw information into meaningful and structured outputs. It involves a series of systematic tasks, such as cleaning, editing, and organizing data, to prepare it for analysis. This step ensures that data errors, duplicates, or inconsistencies are addressed, paving the way for accurate insights.

Effective data processing enhances the usability and reliability of the information, making it ready for interpretation. By processing data correctly, you can uncover patterns, identify trends, and present findings in a way that is easy to understand. This stage serves as a bridge between data collection and analysis, ensuring that the final results are credible and impactful.

Let's delve into the important aspects of data cleaning and editing to understand how they contribute to effective data processing.

9.5.1 Data Cleaning

Data cleaning is the process of identifying and correcting errors, removing duplicates, and ensuring that all data aligns with your research objectives. *For example, if a survey has empty fields for income, you can replace them with the average income or mark them as "unknown."* This step is essential for maintaining the accuracy and quality of your dataset.

Mistakes can happen for various reasons, such as errors during data entry, corrupted files, or mismatched formats when merging datasets. You may also encounter metadata that is incorrectly labeled or inconsistent. To avoid permanent errors, it is always wise to keep a backup of your raw data. This way, if an issue arises during cleaning, you can revert to the original file without losing any important information.

When cleaning data, you can follow the three "C"s:

1. **Complete:** Address missing data by using placeholders or statistical tools to fill in gaps. Ensure that metadata is appropriate for your data type and research topic.

2. **Consistent:** Check that data collected throughout the study is uniform in format, meaning, and scope.

3. **Correct:** Eliminate duplicate records that might lead to inaccurate results and carefully evaluate outliers. Use statistical methods to identify outliers, and decide whether to retain or remove them based on their relevance to your analysis.

9.5.2 Data Editing

Editing may be broadly defined to be a procedure, which uses available information and assumptions to substitute inconsistent values in a data set. In other words, editing is the process of examining the data collected through various methods to detect errors and omissions and correct them for further analysis. *For instance, if survey answers mix measurements (e.g., square feet and square meters), convert all values to one unit for easy comparison.* While editing, care has to be taken to see that the data are as accurate and complete as possible, units of observations and number of decimal places are the same for the same variable.

The following practical guidelines may be handy while editing the data:

1) The editor should have a copy of the instructions given to the interviewers.
2) The editor should not destroy or erase the original entry. Original entries should be crossed out in such a manner that they are still legible.
3) All answers, which are modified or filled in afresh by the editor, have to be indicated.
4) All completed schedules should have the signature of the editor and the date.

For checking the quality of data collected, it is advisable to take a small sample of the questionnaire and examine them thoroughly. This helps in understanding the following types of problems: (1) whether all the questions are answered, (2) whether the answers are properly recorded, (3) whether there is any bias, (4) whether there is any interviewer dishonesty, (5) whether there are inconsistencies. At times, it may be worthwhile to group the same set of questionnaires according to the investigators (whether any particular investigator has specific problems) or according to geographical regions (whether any particular region has specific problems) or according to the sex or background of the investigators, and corrective actions may be taken if any problem is observed.

Before tabulation of data it may be good to prepare an operation manual to decide the process for identifying inconsistencies and errors and also the methods to edit and correct them.

The following broad rules may be helpful:

Incorrect answers: It is quite common to get incorrect answers to many of the questions. A person with a thorough knowledge will be able to notice them. *For example, against the question "Which brand of biscuits do you purchase?" the answer may be "We purchase biscuits from ABC Stores". Now, this questionnaire can be corrected if ABC Stores stocks only one type of biscuits, otherwise not. Answer to the question "How many days did you go shopping in the last week?" would be a number between 0 and 7. A number beyond this range indicates a mistake, and such a mistake cannot be corrected.* The general rule is that changes may be made if one is absolutely sure, otherwise this question should not be used. Usually, a schedule has a number of questions and although answers to a few questions are incorrect, it is advisable to use the other correct information from the schedule rather than discarding the schedule entirely.

Inconsistent answers: When there are inconsistencies in the answers or when there are incomplete or missing answers, the questionnaire should not be used. *Suppose that in a survey, per capita expenditure on various items are reported as follows: Food – Rs. 700, Clothing – Rs.300, Fuel and Light – Rs. 200, other items – Rs. 550 and Total – Rs. 1600.* The answers are obviously inconsistent as the total of individual items of expenditure is exceeding the total expenditure.

Modified answers: Sometimes it may be necessary to modify or qualify the answers. They have to be indicated for reference and checking. Numerical answers to be converted to same units: Against the question "What is the plinth area of your house?" answers could be either in square feet or in square metres. It will be convenient to convert all the answers to these questions in the same unit, square metre for example.

Check Your Progress A:

1. **Fill in the blanks**

   a) _______ is the process of identifying and correcting errors in a dataset to ensure accuracy.

   b) Data editing involves examining the collected data to detect and correct _______ and omissions.

   c) Removing duplicate entries in a dataset is part of _______.

   d) Data cleaning ensures the dataset is _______ and suitable for analysis.

2. **True or False**

   a) Data cleaning includes correcting inconsistent values in the dataset.

   b) Editing data is unnecessary if data cleaning has been performed.

   c) Missing data should always be deleted to avoid errors in analysis.

   d) Outliers in data are automatically corrected during cleaning.

9.6 PRESENTATION OF DATA

Presenting data effectively is an essential step in research and analysis. It transforms raw numbers and facts into a format that is easy to interpret, understand, and communicate. By using various presentation techniques, data can be organized to highlight patterns, trends, and relationships, making it accessible to diverse audiences.

Data can be presented in two primary ways: **tabular presentation** and **visual presentation**.

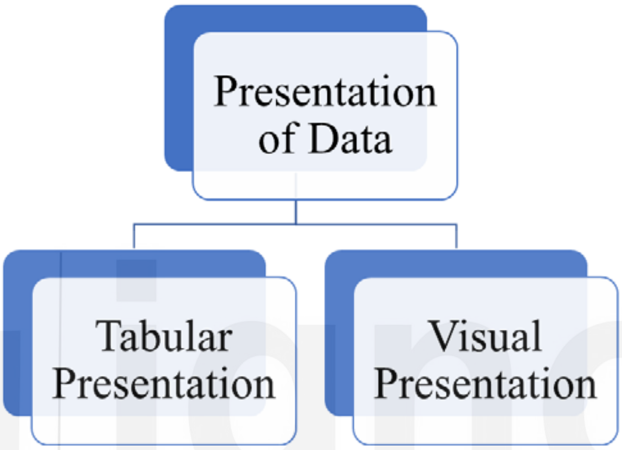


Figure 9.1 Types of Data Presentation

9.6.1 Tabular Presentation of Data

Presentation of collected data in the tabular form is one of the techniques of data presentation. Arranging the data in an orderly manner in rows and columns is called tabulation of data. Sometimes data collected by survey or even from publications of official bodies are so numerous that it is difficult to understand the important features of the data. So, it is important to organize data into tables to make it simple and easy to understand. It may be noted that there may be loss of some minor information in certain cases, but the essential underlying features come out more clearly. Quite frequently, data presented in tabular form is much easier to read and understand than the data presented in the text.

In classification, the data is divided on the basis of similarity and resemblance, whereas tabulation is the process of recording the classified facts in rows and columns. Therefore, after classifying the data into various classes, they should be shown in the tabular form.

Types of Tables

Tables may be classified, depending upon the use and objectives of the data to be presented, into simple tables and complex tables. Let us discuss them along with illustrations.

a) **Simple Table:** In this case, data are presented only for one variable or characteristics. Therefore, this type of table is also known as a one way table. The table showing the data relating to the sales of a company in different years will be an example of a simple table.

Below are the examples of simple type of table:

Illustration 1

Table 9.1: Population of India During 1961–2011 (In thousands)

Census Year	Population
1961	439235
1971	548160
1981	683329
1991	846303
2001	1027015
2011	1210854

Source: Census of India, various documents

Any frequency distribution of a single variable is a simple table

Table 9.2: Frequency Distribution of Daily Wages of 65 Labourers

Daily Wages of Labourers (Rs.)	No. of Labourers
20-30	2
30-40	5
40-50	21
50-60	19
60-70	11
70-80	5
80-90	2
Total	**65**

A simple table may be prepared for descriptive or qualitative data also. The following example illustrates it

Total 9.3: Education of 40 Labourers

Education Level	No. of Persons
Illiterate	22
Literate but below primary	10
Primary	5
High School	2
College and above	1
All	**40**

b) **Complex Table:** A complex table may contain data pertaining to more than one characteristic. The population data given below is an example.

Illustration 2

Table 9.4: Rural and Urban Population of India During 1961–2011 (In thousands)

Census Year	Population		
	Rural	Urban	Total
1961	360298	78937	439235
1971	439046	109114	548160
1981	523867	159463	683329
1991	628691	217611	846303
2001	741660	285355	1027015
2011	833,087	377105	1210854

Note: The total may not add up exactly due to rounding off errors.

Source: Census of India, various documents.

In the above example, the rural and urban population may be subdivided into males and females as indicated below.

Table 9.5: Rural and Urban Population of India During 1961–2011 (sex-wise) (In thousands)

Census Year	Population					
	Rural		Urban		Total	
(1)	(2)	(3)	(4)	(5)	(6)	(7)

In each of the above categories, the persons could be grouped into child and adult, worker and non-worker, or according to different age groups and so on. A particular type of complex table that is of great use in research is a cross-table, where the table is prepared based on the values of two or more variables. The bivariate frequency table is here for illustration.

Illustration 3

Table 9.6: Sales and Profit of 200 Companies

Sl. No.	Sales (Rupees in lakhs)	Profit (Rupees in thousands)						
		Upto 10	10-20	20-50	50-100	100-200	200 an more	Total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	Up to 1	10	3					13
2	1-2	12	12	19				43
3	2-5	11	15	20	10	8		64
4	5-10	2	8	15	5	10		40
5	10-20		2	12	4	9	6	33
6	20 and more			2	1	2	2	7
7	Total	35	40	68	20	29	8	200

From a bivariate table, one may get some idea about the interrelationship between two variables. Suppose that all the frequencies are concentrated in the diagonal cells, then there is likely to be a strong relationship. That is a positive relationship if it starts from top-left corner to bottom-right corner or if it is from bottom-left corner to top-right corner then, we could say there is negative relationship. If the frequencies are more or less equally distributed over all the cells, then probably there is no strong relationship.

Multivariate tables may also be constructed but interpretation becomes difficult once we go beyond two variables.

So far, we have discussed and learnt about the types of tables and their usefulness in presentation of data. Now, let us proceed to learn about the different parts of a table, which enable us to have a clear understanding of the rules and practices followed in the construction of a table.

Parts of A Statistical Table

A table should have the following four essential parts - title, caption or box head (column), stub (row heading) and main data. At times it may also contain an endnote and source note below the table. The table should have a title, which is usually placed above the statistical table. The title should be clearly worded to give some idea of the table's contents. Usually, a report has many tables. Hence the tables should be numbered to facilitate reference.

Caption refers to the title of the columns. It is also termed as "box head". There may be sub-captions under the main caption. Stub refers to the titles given to the rows.

Caption and stub should also be unambiguous. To the extent possible abbreviations should not be used in either caption or stub. But if they are used, the expansion must be given in the end note below. Notes pertaining to stub entries or box headings may be numerals. But, to avoid confusion, it is better to use some symbols (like *, **, @ etc.) or alphabets for notes referring to the entries in the main body. If the table is based on outside information, it should be mentioned in the source note below. This note should be complete with author, title, year of publication etc. to enable the reader to go to the original source for cross checking or for obtaining additional information. Columns and rows may be numbered for easy reference.

Some of these features are illustrated below with reference to the table on Rural and Urban Population during 1961-2001, which was presented in earlier Illustration-2, Table 9.4.

1. Title of the Table	Table: Rural and Urban Population of India during 1961-2001 (in thousands)

2. Caption or Box head	**Population**		
	Rural	Urban	Total

3. Stub (Row Heading)	Census Year
	1961
	1971
	1981
	1991
	2001

4. Body (Main Data)	360298	78937	439235
	439046	109114	548160
	523867	159463	683329
	628691	217611	846303
	741660	285355	1027015

5. End Note	Note: The total may not add up exactly due to rounding off of error.

6. Source Note	Source: Census of India, various documents.

Column Number	(1)	(2)	(3)	(4)	(5)

Row Number	1
	2
	3
	4
	5

The boxes above are self-explanatory.

Arrangement of items in stub and box-head

There is no hard and fast rule about the arrangement of column and row headings in a table. It depends on the nature of data and type of analysis. A number of different methods are used - alphabetical, geographical, chronological/historical, magnitude-based and customary or conventional.

Alphabetical: This method is suitable for general tables as it is easy to locate an item if it is arranged alphabetically. *For example, population census data of India may be arranged in the alphabetical order of states/union territories.*

Geographical: It can be used when the reader is familiar with the usual geographical classification.

Chronological: A table containing data over a period of time may be presented in the chronological order. Population data (1961 to 2001) presented earlier (Tables 9.1 and 9.4) are in chronological order. One may either start from the most recent year or the earliest year. However, there is a convention to start with the month of January whenever year and month data are presented.

Based on Magnitude: At times, items in a table are arranged according to the value of the characteristic. Usually, the largest item is placed first and other items follow in decreasing order. But this may be reversed also. Suppose that state-wise population data is arranged in order of decreasing magnitude. This will highlight the most populous state and the least populous state.

Customary or Conventional: Traditionally some order is followed in certain cases. While presenting population census data, usually 'rural' comes before 'urban' and 'male' first and 'female' next. At times, conventional geographical order is also followed.

One point may be noted. The above arrangements are not exclusive. In a big table, it is always possible and sometimes convenient to arrange the items following two or three methods together. For example, it is possible to construct a table in chronological order and within it in geographical order. Sometimes information from the same table may be rearranged to produce another table to highlight certain aspects. This will be clear from the following specimen tables.

Table A

Sl.No.	**Census Year**	**Rural**		**Urban**	
		Male	**Female**	**Male**	**Female**
(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
1					
2					

Table B

Sl.No.	**Census Year**	**Male**		**Female**	
		Rural	**Urban**	**Rural**	**Urban**
(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
1					
2					

Tables A and B contain the same information. Table A compares male-female differences for rural and urban areas whereas Table B highlights rural-urban contrasts for both the sexes.

Tables are prepared for making data easy to understand for the reader. It should not be very large as the focus may be lost. A large table may be logically broken into two or more small tables.

Requisites of a Good Statistical Table

After having an understanding of the parts of a statistical table, now let us discuss the features of an ideal statistical table. Besides the rules relating to part of the table, certain guidelines are very helpful in its preparation. They are as follows:

1) A good table must present the data in as clear and simple a manner as possible.

2) The title should be brief and self-explanatory. It should represent the description of the contents of the table.
3) Rows and Columns may be numbered to facilitate easy reference.
4) Table should not be too narrow or too wide. The space of columns and rows should be carefully planned, so as to avoid unnecessary gaps.
5) Columns and rows which are directly comparable with one another should be placed side by side.
6) Units of measurement should be clearly shown.
7) All the column figures should be properly aligned. Decimal points and plus or minus signs also should be in perfect alignment.
8) Abbreviations should be avoided in a table. If it is inevitable to use, their meanings must be clearly explained in the footnote.
9) If necessary, the derived data (percentages, indices, ratios, etc.) may also be incorporated in the tables.
10) The sources of the data should be clearly stated so that the reliability of the data could be verified, if needed.

Check Your Progress B

The following report is obtained from 50 unskilled workers in a factory in Faridabad. Prepare three simple tables based on caste, education and place of origin and a complex table by considering all the factors.

S.No.	Caste	Education	Place of Orign
1	OC	Literate but below primary	Urban
2	OC	Literate but below primary	Urban
3	OC	Primary	Rural
4	BC	Illiterate	Rural
5	SC	Literate but below primary	Urban
6	SC	Primary	Rural
7	ST	Illiterate	Rural
8	OC	High School	Rural
9	SC	High School	Rural
10	ST	Illiterate	Rural
11	OC	Literate but below primary	Rural
12	OC	Primary	Rural
13	OC	High School	Rural
14	BC	Illiterate	Rural
15	SC	Primary	Rural
16	SC	Primary	Rural
17	OC	High School	Rural
18	OC	High School	Rural
19	SC	Literate but below primary	Rural
20	ST	Primary	Urban
21	SC	Primary	Rural

22	SC	High School	Rural
23	ST	Primary	Rural
24	OC	Literate but below primary	Rural
25	OC	Primary	Rural
26	OC	Primary	Rural
27	OC	Primary	Urban
28	OC	Primary	Urban
29	BC	Illiterate	Rural
30	SC	Illiterate	Rural
31	SC	Illiterate	Urban
32	ST	Illiterate	Urban
33	OC	Primary	Urban
34	OC	Primary	Rural
35	BC	Illiterate	Rural
36	OC	High School	Urban
37	OC	High School	Urban
38	OC	Illiterate	Rural
39	BC	Illiterate	Urban
40	OC	Primary	Rural
41	BC	Literate but below primary	Rural
42	BC	Illiterate	Urban
43	OC	Illiterate	Rural
44	BC	Primary	Rural
45	BC	Literate but below primary	Rural
46	SC	Primary	Rural
47	SC	Illiterate	Urban
48	ST	Primary	Rural
49	OC	Primary	Rural
50	OC	Literate but below primary	Rural

9.6.2 Visual Presentation of Data

Visual presentation of statistical data has become more popular and is often used by the researcher and the statistician in analysis. Visual presentation of data means presentation of Statistical data in the form of diagrams and graphs. These days, as we know, every research work is supported with visual presentation because of the following reasons:

1) **They relieve the dullness of the numerical data:** Any list of figures becomes less comprehensible and difficult to draw conclusions from as its length increases. Scanning of the figures from tables causes undue strain on the mind. The data, when presented in the form of diagrams and graphs, gives a bird's eye-view of the entire data and creates interest and leaves an impression on the mind of readers for a long period.

2) **They make comparison easy:** This is one of the prime objectives of visual presentation of data. Diagrams and graphs make quick comparison

between two or more sets of data simpler, and the direction of curves bring out hidden facts and associations of the statistical data.

3) **They save time and effort:** The characteristics of statistical data, through tables, can be grasped only after a great strain on the mind. Diagrams and graphs reduce the strain and save a lot of time in understanding the basic characteristics of the data.

4) **They facilitate the location of various statistical measures and establish trends:** Graph makes it possible to locate several measures of central tendency such as Median, Quartiles, Mode etc. They help in establishing trends of the past performance and are useful in interpolation or extrapolation, line of best fit, establishing correlation etc. Thus, it helps in forecasting.

5) **They have universal applicability:** It is a universal practice to present numerical data in the form of diagrams and graphs. These days, it is an extensively used technique in the field of economics, business, education, health, agriculture etc.

6) **They have become an integral part of research:** In fact, nowadays it is difficult to find any research work without visual support. The reason is that this is the most convincing and appealing way of presenting the data. You can find diagrammatic and graphical presentation of data in journals, magazines, television, reports, advertisements etc. After having understood about the importance of visual presentation, we shall move on to discuss the Diagrams and graphs which are more frequently used in the area of business research.

Here are some common types of visual data presentation methods

9.6.2.1 Line Graph

A line graph is a powerful tool used to represent data trends over time or continuous variables. It consists of points plotted on a graph that are connected by a line, making it easy to observe increases, decreases, or fluctuations in data. Line graphs are especially useful for showing relationships and patterns over time, such as growth, decline, or stability.

Characteristics of a Line Graph:

- **X-axis**: Represents the independent variable, usually time or categories (e.g., months, years).
- **Y-axis**: Represents the dependent variable, such as quantity, percentage, or value.
- Points are plotted and connected to form a continuous line.

Example:

Suppose a business tracks its monthly revenue for six months. The data is as follows:

- January: Rs.10,000
- February: Rs.12,000

- March: Rs.11,000
- April: Rs.14,000
- May: Rs.13,000
- June: Rs.15,000

When this data is represented on a line graph, the months are placed on the X-axis, and revenue figures are on the Y-axis. Connecting the points for each month forms a line, clearly showing how the revenue increased, dipped slightly, and then rose again.

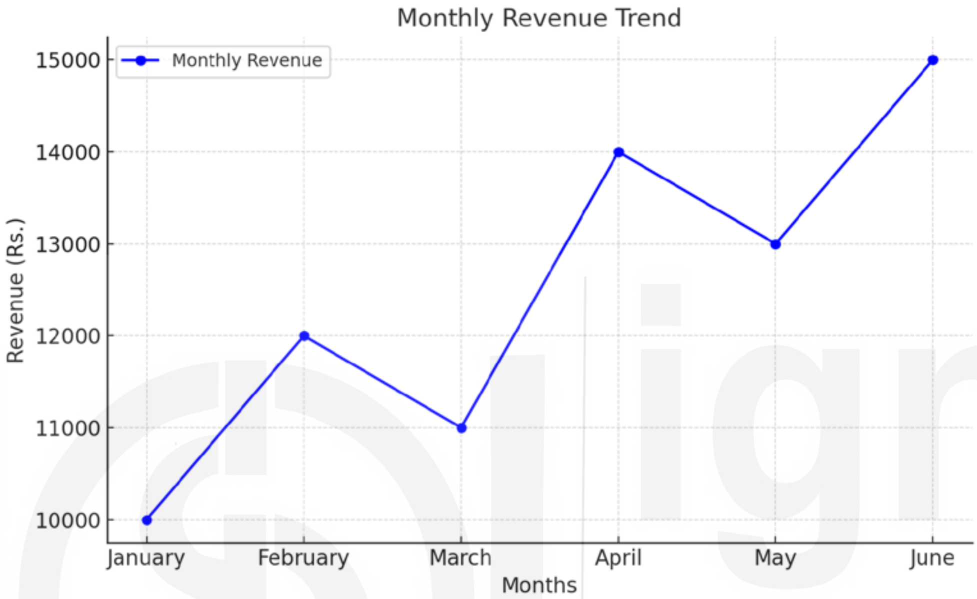


Figure 9.2: Representation of Line Graph

Benefits of Using Line Graphs:

- They make trends and changes easy to identify.
- Multiple datasets can be compared by using different lines on the same graph.
- They provide a clear visual representation of fluctuations over time.

Line graphs are widely used in business, economics, and research to monitor performance, predict future outcomes, and communicate data effectively.

9.6.2.2 Histogram

The graph usually drawn to represent a frequency distribution is called a Histogram. A histogram is a set of rectangles (vertical bars) each proportionate in width to the magnitude of a class interval and proportionate in area to the number of frequencies concerning the classes' intervals. In a histogram, the variables (class-intervals) are always shown on the X-axis and the frequencies are taken on the Y-axis.

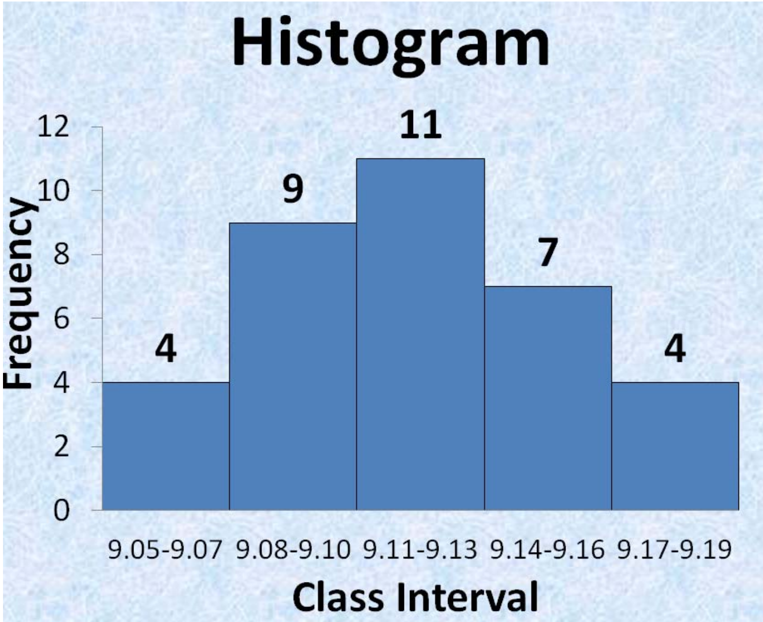


Figure 9.3 Representation of Histogram

In constructing a histogram there should not be any gap between two successive rectangles, and the data must be in the exclusive form of classes. However, we cannot construct a histogram for distribution with open-end classes and it can be quite misleading if the distribution has unequal class intervals.

The value of mode can be determined from the histogram. The procedure for locating the mode is to draw a straight line from the top right corner of the highest rectangle (Modal Class) to the top right corner of the preceding rectangle (Pre Modal Class). Similarly, draw a straight line from the top left corner of the highest rectangle to top left corner of the succeeding rectangle (Post Modal Class). Draw a perpendicular from the point of intersection of these two straight lines to the X-axis. The point where it meets the X-axis gives the value of mode. This is shown in Figure 9.4. However, graphic location of Mode is not possible in a multi-distribution.

9.6.2.3 Frequency Polygon

A frequency polygon is a type of line graph used to represent frequency distribution. In this graph, the class frequency is plotted against the class midpoint, and the points are connected by straight lines to form a curve. It helps highlight the highs and lows in the data and is useful for comparing two or more frequency distributions. The term "polygon" means "many-angled diagram," and a frequency polygon provides a clear and simple way to visualize data. It can be drawn directly from the data or based on a histogram.

The procedure for the construction of a frequency polygon by histogram is to first draw the histogram, as explained earlier, of the given data. Then, put a dot at the midpoint of the top horizontal line of each rectangle bar and join these dots by straight lines.

Another way of drawing frequency polygons is to obtain the mid-values of class intervals and plot them on the X-axis. Mark frequency along the Y axis. Then, plot the frequency values corresponding to each midpoint and connect

them through straight lines. The area left outside is just equal to the area included in it. Hence, the area of a polygon is equal to the area of histogram. The difference between the histogram and the polygon is that the histogram depicts the frequency of each class separately whereas the polygon does it collectively. The histogram is usually associated with the data of discrete series, while frequency polygon is for continuous series data.

Let us, now, take up an illustration to learn how to draw a histogram, and frequency polygon practically and also determine the mode. The data relates to the sales of computers by different companies.

Illustration-4

Salex (Rs. In crores)	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of Companies	8	20	35	50	90	70	30	15

Solution : For drawing histogram, as explained earlier, we have to show sales on X - axis and number of companies on Y-axis by selecting a suitable scale. For drawing frequency polygons, plot dots on the top middle of each rectangle, and join them by straight lines.

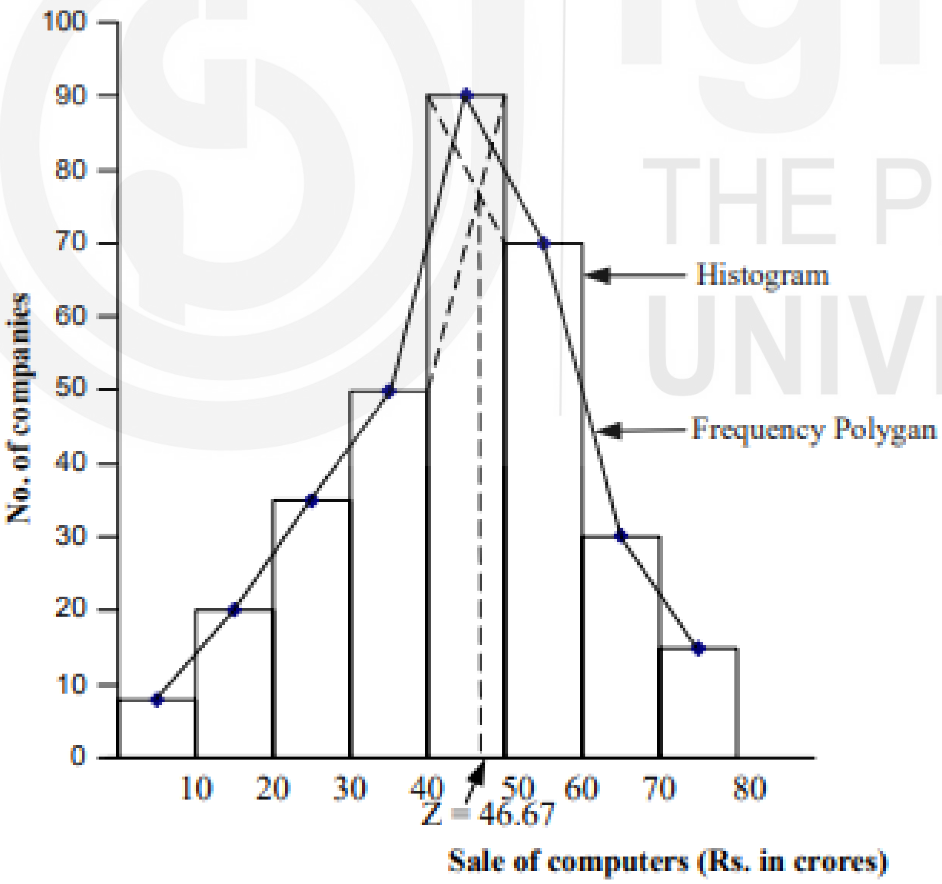


Figure 9.4: Histogram and Frequency Polygon

Self Assessment Exercise

The monthly production of units by a sample of 200 workers in a bulbs manufacturing firm is given in the following table.

Output (Units)	200-225	225-250	250-275	275-300	300-325	325-350	350-375	375-400
No. of Workers	12	21	25	40	49	28	17	8

i) Draw a histogram and frequency polygon.

ii) Determine the mode graphically.

9.6.2.4 Bar Diagram

Bar graphs are the pictorial representation of data (generally grouped), in the form of vertical or horizontal rectangular bars, where the length of bars are proportional to the measure of data. They are also known as bar charts.

Bar refers to a thick line. Under this type of construction only one dimension i.e., length is taken into account for the purpose of comparison and observance of fluctuations in growth. The length of each bar is proportionate to the magnitude of the data. The width is not related to the magnitude of the data. Generally the width is given for the purpose of visual effect and attractiveness. The width of each bar and the gap between one bar to another bar must be uniform. Mention the respective figures at the top of every bar, particularly when the scale is too narrow, so that the reader knows the figures without consulting the scale of the diagram.

A large number of one dimensional diagrams are available for presenting data. Such as line diagram, simple bar diagram, multiple bar diagram, sub-divided bar diagram, percentage bar diagram, deviation bar diagram etc. We shall, however, study only the simple bar diagram, multiple bar diagram, and sub-divided bar diagram. Let us study these three kinds of diagrams with the support of relevant illustrations.

Simple Bar Diagram

In a Simple bar diagram, the data related to one variable is depicted. Such as, profits, investments, exports, sales, production etc.

This type of diagram may be drawn either vertically or horizontally. Both positive and negative values can be presented. In such a case, if bars are constructed vertically, the positive values are taken on the upper side of the horizontal axis while the negative values are taken on its lower side. On the other hand if the bars are constructed horizontally, the positive values are taken on the right hand side of the vertical axis and the negative values are considered on its left side. These type of construction of bars are also called deviation bar diagrams. The simple bar diagram is very easy to prepare and to understand the level of fluctuations from one situation to another. It should be kept in mind that only length is taken into account and not width. Width should be uniform for all bars and the gap between each bar is normally

identical. Let us consider the following illustrations and learn how to present the given data in the form of simple bar diagrams vertically and horizontally.

Illustration-5

Prepare a Simple Bar Diagram from the Following Data Relating to Tea Exports

Year	**2018-19**	**2019-20**	**2020-21**	**2021-22**	**2022-23**	**2023-24**	**2024-25**
Exports (In Million kgs.)							

Solution: The quantity of tea exported is given in million kgs. for different years. A simple bar diagram will be constructed with 7 bars corresponding to the 7 years. Now study the following vertical construction of bar diagrams by referring to the guidelines for construction of simple bars, as explained in section7.6.2.4

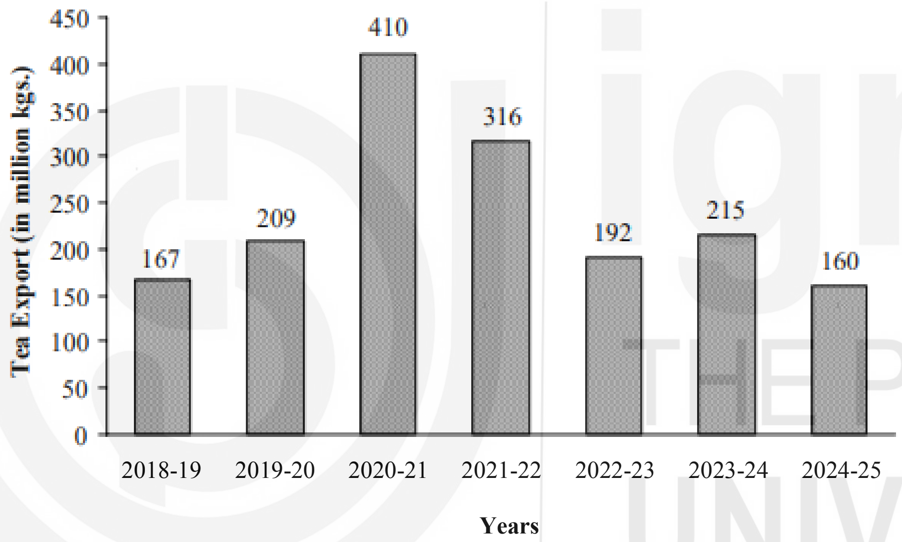


Figure 9.5: Simple Bar Diagram Showing the Tea Exports in Different Years.

Illustration-6

The following data relates to the Profit and Loss of different industries in 1999- 2002. Present the data through a simple bar diagram.

Industry:	**Cement**	**Oil**	**Textile**	**Sugar**	**Garments**
Exports (In Million kgs.)					

Solution: The given data represents positive and negative values i.e., profit and loss. Let us draw the bars horizontally. Observe Figure 9.6 carefully and try to understand the construction of simple bars horizontally

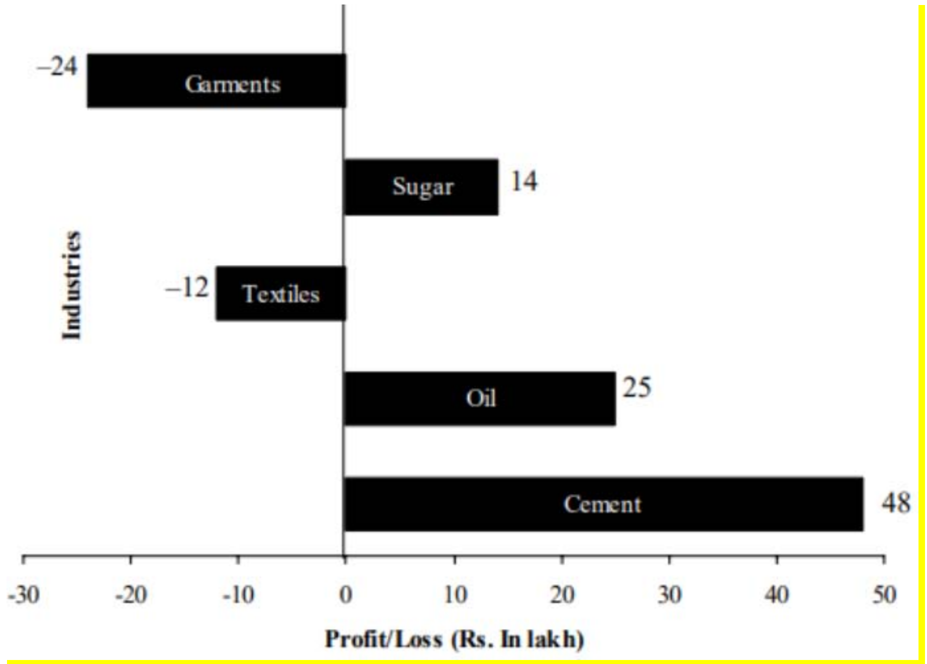


Figure 9.6: Simple Bar Diagram Showing the profit and Loss of Different Industries During 1992-02

Check Your Progress C

Represent the following data related to the surplus/deficit of Balance of Trade over a period, by simple bar diagram.

Years	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**
Surplus(+)deficit(–) (*In million $*)							

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

9.6.2.5 Pie Chart

A Pie Chart is a circular statistical graphic that is divided into slices to illustrate numerical proportions. Each slice represents a category of data, and the size of the slice corresponds to the proportion or percentage it holds in the total. Pie charts are particularly useful for showing relative sizes or comparisons between parts of a whole.

Characteristics of a Pie Chart

1. **Circular Format:** The chart is a complete circle, representing 100% of the data.
2. **Proportional Slices:** The size of each slice is proportional to the value it represents.
3. **Labeling:** Each slice is usually labeled with its category and percentage or value.
4. **Color Coding:** Different colors or patterns are often used to distinguish between slices for better visualization.

Steps to Create a Pie Chart

1. **Collect Data:** Gather data that can be categorized and expressed as percentages of a whole.
2. **Calculate Proportions:** Determine the percentage share of each category relative to the total.
3. **Divide the Circle:** Allocate slices based on the calculated proportions (e.g., a 25% category takes up a quarter of the circle or 90°).
4. **Label the Slices:** Assign clear labels, including category names and percentages, to make the chart interpretable.
5. **Use Colors or Patterns:** Apply distinct colors or patterns for clarity and visual appeal.

Example of a Pie Chart

Suppose a company's total sales revenue of Rs.1,00,000 is distributed across four product categories:

- Product A: Rs. 40,000 (40%)
- Product B: Rs. 25,000 (25%)
- Product C: Rs. 20,000 (20%)
- Product D: Rs. 15,000 (15%)

This data can be represented as a pie chart, with each slice corresponding to the respective percentage of total sales.

Product	Sales (Rs)	Percentage	Angle in Pie Chart
Product A	40,000	40%	(40/100) × 360 = 144°
Product B	25,000	25%	(25/100) × 360 = 90°
Product C	20,000	20%	(20/100) × 360 = 72°
Product D	15,000	15%	(15/100) × 360 = 54°

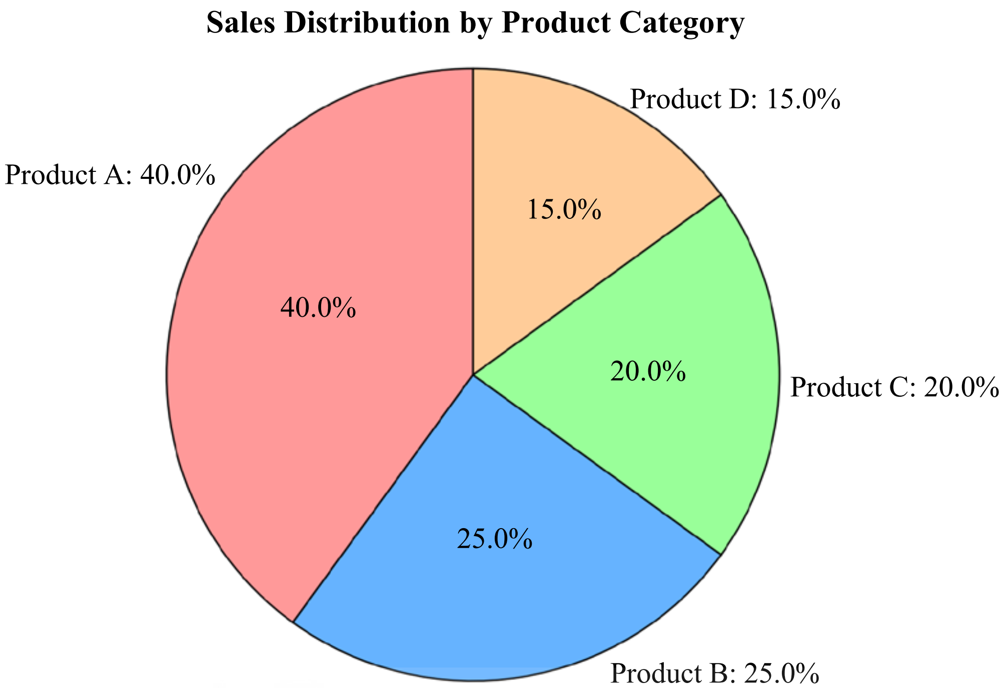


Figure 9.7 Pie Chart

Advantages of Pie Charts

1. **Simple Visualization:** Ideal for a quick understanding of proportions.
2. **Clear Representation:** Shows the relationship between parts and the whole effectively.
3. **Appealing Presentation:** Visually engaging for non-technical audiences.

Limitations of Pie Charts

1. **Limited Categories:** Best suited for datasets with 5-6 categories; too many slices can make it cluttered.
2. **Lack of Precision:** Difficult to interpret exact values or compare close percentages.
3. **Not Suitable for Time Trends:** Pie charts are static and cannot show changes over time.

When to Use a Pie Chart

- To show the proportionate distribution of a single dataset (e.g., market share, budget allocation).
- To compare parts of a whole rather than detailed numerical data.
- When the audience is more likely to benefit from a visual summary than exact figures.

9.7 LET US SUM UP

In this unit, we explored the critical aspects of data preparation and presentation, which are essential steps in research and analysis. Data preparation serves as the foundation for reliable and accurate results. It involves cleaning, editing, and organizing raw data to ensure it is free from errors, inconsistencies, or missing values that could otherwise lead to flawed conclusions. For instance, removing duplicate entries, correcting incorrect

values, and addressing gaps in a dataset are vital for producing trustworthy insights. Data editing further refines this process by standardizing units, verifying accuracy, and maintaining uniformity across datasets. Together, these steps enhance the overall quality and usability of data, making it ready for meaningful analysis.

Data presentation, on the other hand, focuses on transforming this prepared data into a format that effectively communicates findings. It bridges the gap between analysis and decision-making by presenting information in a clear, concise, and visually appealing manner. Tabular presentation, which organizes data into rows and columns, is particularly effective for detailed comparisons or when precision is required. For example, a table showing population trends over decades allows researchers to identify changes at a glance. Visual presentation, through tools such as line graphs, histograms, bar diagrams, and pie charts, simplifies complex datasets and highlights trends or patterns that might not be immediately apparent in raw numbers. These visual aids make it easier for audiences to grasp key insights and support informed decision-making.

By combining effective data preparation and presentation, researchers can ensure their findings are not only accurate but also impactful. These skills are indispensable for anyone engaged in research, enabling them to handle data responsibly and communicate results in a way that resonates with diverse audiences. Mastery of these techniques empowers researchers to convert raw information into actionable knowledge, ultimately contributing to better outcomes in research, business, and beyond.

9.8 KEYWORDS

Bar Diagram: A chart with rectangular bars representing grouped data, with the length of the bars proportional to the data values.

Categorical Data: Data divided into distinct categories, often represented in pie charts or bar diagrams.

Data Cleaning: The identification and correction of errors, removal of duplicate entries, and handling of missing values to ensure data accuracy.

Data Editing: The systematic process of reviewing data to detect inconsistencies and correcting them for further analysis.

Data Preparation: The process of cleaning, organizing, and structuring raw data to make it suitable for analysis.

Frequency Polygon: A type of line graph that shows frequencies by connecting midpoints of histogram bars.

Histogram: A bar chart used to represent the frequency distribution of data, where bars are proportional to class intervals.

Line Graph: A graphical representation of data trends over time, using points connected by a continuous line.

Metadata: Information about data that provides context and ensures its usability and relevance.

Numerical Data: Quantitative data that can be represented using histograms, line graphs, or frequency polygons.

Outliers: Data points that significantly differ from the rest of the dataset, which may need evaluation for their relevance.

Pie Chart: A circular chart divided into slices to illustrate numerical proportions or percentages of a dataset.

Tabular Presentation: Organizing data into rows and columns for easy interpretation and analysis.

Visual Presentation: Representing data graphically using tools such as line graphs, histograms, bar diagrams, and pie charts.

9.9 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress A

1. **Answers to Fill in the Blanks:**

   a) Data cleaning b) Errors c) Data cleaning d) Accurate

2. **Answers to True/False:**

   a) True b) False c) False d) False

9.10 TERMINAL QUESTIONS

1. Explain the significance of data preparation in research.
2. Discuss the key steps involved in the processing of data. How do data cleaning and data editing contribute to data accuracy?
3. What is the role of tabular presentation in data presentation? Provide an example where tabular presentation is more effective than visual presentation.
4. Describe the advantages and disadvantages of using visual data presentation methods such as bar diagrams, pie charts, and line graphs.
5. Compare and contrast a histogram and a frequency polygon. In which scenarios are each of these most effective?
6. How would you choose the most appropriate data presentation method for a given dataset? Provide criteria to evaluate the suitability of tabular and visual methods.

9.11 FURTHER READINGS

1. Black, S. C. (1991). Data analysis and presentation. In *Instrumental analysis of pollutants* (pp. 335-355). Dordrecht: Springer Netherlands.
2. In, J., & Lee, S. (2017). Statistical data presentation. *Korean journal of anesthesiology*, *70*(3), 267-276.
3. Kothari, C. R. (2004). Research methodology.
4. Mittal, S., Chawla, D., & Sondhi, N. Journal of Indian Business Research.

UNIT 10 DESCRIPTIVE STATISTICS

Structure

10.0 OBJECTIVES

After studying this unit, you will be able to:

- Understand the fundamental concepts and significance of descriptive statistics in data analysis;
- Compute measures of central tendency mean, median, and mode for summarizing data;
- Understand and apply measures of dispersion range, standard deviation, and coefficient of variation; and
- Analyze relationships between mean, median, and mode and their application in interpreting data distributions.

10.1 INTRODUCTION

In the previous unit, we explored the processes of data preparation and presentation, focusing on organizing raw data into structured formats such as tables, charts, and graphs. This process ensures that datasets are clean, organized, and visually comprehensible. Building upon that foundation, this unit explores into descriptive statistics, which further simplifies and summarizes the data by extracting key numerical insights.

Descriptive statistics are the foundation of statistical analysis. They help us organize and summarize data so that it is easier to understand. They are often the first step in analysing data, providing a clear way to make sense of raw information. In simple terms, descriptive statistics turn large amounts of unorganized data into meaningful insights, helping people recognize patterns, relationships, and trends within the data.

For example, imagine a retail company looking at its monthly sales. Without descriptive statistics, it would be very hard to find any trends or understand customer behaviour from thousands of individual sales transactions. By calculating the average monthly sales (mean) and looking at how sales numbers vary (range or standard deviation), the company can figure out not only the typical performance but also how much sales differ from month to month. This information could help the company plan better, improve resources, and focus on the months that need extra attention.

Descriptive statistics are useful because they provide a simple summary of the data, including common values (like the mean, median, and mode) and variability (like range and standard deviation). By using charts, graphs, and simple numbers, descriptive statistics make complicated information easier to understand and use.

With the help of descriptive statistics, messy and overwhelming datasets become organized and easy to work with. Thus clear understanding of the data is the foundation for deeper analysis and better decision-making in many areas, including business, healthcare, education, and engineering.

10.2 SIGNIFICANCE OF DESCRIPTIVE STATISTICS

Descriptive statistics are invaluable for analyzing and interpreting data effectively. Their importance can be understood through various real-life examples and applications:

Data Simplification:

Descriptive statistics condense large datasets into easy-to-understand summaries. *For example, instead of looking at thousands of individual sales figures, a company can focus on key metrics like the average sales (mean), the most frequent sales figure (mode), or the range of sales.* We will discuss this concept mean, mode, median, range, and standard deviation further in this unit. This simplification helps stakeholders quickly grasp the overall performance without diving into granular details.

Facilitate Comparisons:

By summarizing data, descriptive statistics enable comparisons across different groups or time periods. *For example, a healthcare organization can compare the average patient wait times across different departments or shifts, helping identify which area needs improvement.*

Identify Trends and Patterns:

Descriptive statistics help highlight trends and patterns within the data. *For example, a school can analyze students' exam scores over time to identify whether performance is improving, declining, or remaining consistent.* Patterns such as peak performance during specific months can inform future planning.

Data Quality Assessment:

Measures like standard deviation and range help assess the variability of data. *If a manufacturer finds a high standard deviation in the sizes of a product batch, it may indicate issues in the production process that require correction.* Similarly, identifying outliers can flag potential errors or special cases needing further investigation.

Support Evidence-Based Decisions:

In the business world, descriptive statistics form the foundation of strategic decisions. *For example, a retailer analyzing customer purchase behavior might use the average transaction value and customer frequency to design loyalty programs.* In healthcare, hospitals can use descriptive statistics to optimize staff allocation based on patient influx patterns.

Visualization and Communication:

Descriptive statistics are often presented through graphs and charts, making it easier to communicate findings to non-technical stakeholders. *For example, a pie chart showing the percentage of sales by product category can help management focus on the most profitable areas.*

Predictive Insights:

While descriptive statistics do not predict future trends directly, they provide a foundation for more advanced analyses. *For example, understanding seasonal sales trends can help a retailer stock inventory appropriately in anticipation of peak demand.*

Examples of Applications of Descriptive Statistics in our day to day life:

Education: Schools use descriptive statistics to summarize test scores. For instance, a teacher can calculate the class average (mean) and identify the highest and lowest scores (range) to assess overall class performance.

Healthcare: Hospitals analyze patient data using descriptive statistics. For example, calculating the average time spent in the emergency room helps in resource planning.

Finance: Investment firms use measures like average return and standard deviation to assess the performance and risk of financial portfolios.

Marketing: Businesses analyze customer data, such as average purchase value and most frequently purchased items, to design targeted campaigns.

Cricket: Teams use descriptive statistics to assess players' performances. For example, calculating the batting average and strike rate of batters or evaluating bowlers based on their economy rate and wicket averages helps in selecting the best players for specific matches or formats.

By providing simplified and actionable insights, descriptive statistics make data-driven decision-making accessible and reliable across industries and domains.

10.3 MEASURES OF CENTRAL TENDENCY

Measures of central tendency i.e. condensing the mass of data in one single value, enable us to get an idea of the entire data. For example, it is impossible to remember the individual incomes of millions of earning people of India. But if the average income is obtained, we get one single value that represents the entire population. Measures of central tendency also enable us to compare two or more sets of data to facilitate comparison. *For example, the average sales figures of April may be compared with the sales figures of previous months.* A good measure of central tendency must possess the following characteristics:

It should be rigidly defined: The definition of a measure of central tendency should be clear and unambiguous so that it leads to one and only one information.

It should be based on all observations: A good measure of central tendency should be based on all the values of the distribution of scores.

It should be amenable for further mathematical treatment: If we are given two sets of data and a measure of central tendency for both of them, we should be able to calculate the measure for the combined data also.

It should be least affected by the fluctuation of sampling: If independent random samples of the same size are selected from a population, the value of average for each one of them should be sufficiently close to one another.

10.3.1 Arithmetic Mean

The arithmetic mean (or mean or average) is the most commonly used and readily understood measure of central tendency. In statistics, the term average refers to any of the measures of central tendency. The arithmetic mean is defined as being equal to the sum of the numerical values of each and every observation divided by the total number of observations. Symbolically, it can be represented as:

$$\bar{x} = \frac{\sum fx}{N}$$

where ∑x indicates the sum of the values of all the observations, and N is the total number of observations. For example, let us consider the monthly salary (Rs.) of 10 employees of a firm

2500, 2700, 2400, 2300, 2550, 2650, 2750, 2450, 2600, 2400

If we compute the arithmetic mean, then

$$\bar{X} = 2500 + 2700 + 2400 + 2300 + 2550 + 2650 + 2750 + 2450 + 2600 + 2400$$

$$= \frac{25300}{10} = Rs. 2530.$$

Therefore, the average monthly salary is Rs. 2530.

We have seen how to compute the arithmetic mean for ungrouped data. Now let us consider what modifications are necessary for grouped data. When the observations are classified into a frequency distribution, the midpoint of the class interval would be treated as the representative average value of that class. Therefore, for grouped data; the arithmetic mean is defined as

$$\bar{x} = \frac{\sum fx}{N}$$

Where X is midpoint of various classes, f is the frequency for corresponding class and N is the total frequency, i.e. N = $\sum$f.

This method is illustrated for the following data which relate to the monthly sales of 200 firms.

Table 10.1 Monthly sales of 200 firms

Monthly Sales (Rs. Thousand)	No. of Firms	Monthly Sales (Rs. Thousand)	No. of Firms
300-350	5	550-600	25
350-400	14	600-650	22
400-450	23	650-700	7
450-500	50	700-750	2
500-550	52		

For computation of arithmetic mean refer table 10.2:

Table 10.2: Monthly Sales of 200 Firms

Monthly Sales (Rs. Thousand)	Mid point X	No. of Firms f	fX
300-350	325	5	1625
350-400	375	14	5250
400-450	425	23	9775
450-500	475	50	23750
500-550	525	52	27300
550-600	575	25	14375
600-650	625	22	13750
650-700	675	7	4725
700-750	725	2	1450
		N = 200	**$\sum$fX = 102000**

$$\bar{X} = \frac{\sum fX}{N} = \frac{102000}{200} = 510$$

Hence the average monthly sales are Rs. 510.

To simplify calculations, the following formula for arithmetic mean may be more convenient to use.

It may be observed that this formula is much faster than the previous one and the value of arithmetic mean remains the same.

$$\bar{x} = A + \frac{\sum fd}{N} \times i$$

where A is an arbitrary point, $d = \frac{X-A}{i}$, and i = size of the equal class interval.

REMARK: A justification of this formula is as follows. When $d = \frac{X-A}{i}$, then $X = A + i\ d$ Multiplying throughout by f, taking summation on both sides and. Dividing by N, we get

$$\bar{x} = A + \frac{\sum fd}{N} \times i$$

This formula makes the computations very simple and takes less time. To apply this formula, let us consider the same example discussed earlier and shown again in the table 10.3.

Table 10.3: Average Sales Calculation through Assumed (Arbitrary) Mean

Monthly Sales (Rs. Thousand)	**Mid Point**	**No. of Firms f**	**(x-525)/50 = d**	**fd**
300-350	325	5	-4	-20
350-400	375	14	-3	-42
400-450	425	23	-2	-46
450-500	475	50	-1	-50
500-550	525	52	0	0
550-600	575	25	+1	+25
600-650	625	22	+2	+44
650-700	675	7	+3	+21
700-750	725	2	+4	+8
			N = 200	**∑fd = -60**

$$\bar{X} = A + \frac{\sum fd}{N} \times i = 525 - \frac{60}{200} \times 50$$

$$= 525-15 = 510 \text{ or Rs. } 510$$

It may be observed that this formula is much faster than the previous one and the value of arithmetic mean remains the same.

Advantages of Mean

1) The definition of mean is rigid which is a quality of a good measure of central tendency.
2) It is not only easy to understand but also easy to calculate.
3) All the scores in the distribution are considered when the mean is computed.
4) Further mathematical calculations can be carried out on the basis of mean.
5) Fluctuations in sampling are least likely to affect the mean.

Limitations of Mean

1) Outliers or extreme values can have an impact on mean.
2) When there are open ended classes, such as 10 and above or below 5, mean cannot be computed. In such cases median and mode can be computed. This is mainly because in such distributions mid-point cannot be determined to carry out calculations.
3) If a score in the data is missing or lost or not clear, then mean cannot be computed unless mean is computed for the rest of the data by not considering the lost score and dropping it all together.
4) It is not possible to determine the mean through inspection. Further, it cannot be determined based on a graph.
5) It is not suitable for data that is skewed or is very asymmetrical as then in such cases the mean will not adequately represent the data.

10.3.2 Median

A second measure of central tendency is the median. Median is that value which divides the distribution into two equal parts Fifty per cent of the observations in the distribution are above the value of median and other fifty per cent of the observations are below this value of median. The median is the value of the middle observation when the series is arranged in order of size or magnitude. If the number of observations is odd, then the median is equal to one of the original observations. If the number of observations is even, then the median is the arithmetic mean of the two middle observations. For example, if the income of seven persons in rupees is 1100, 1200, 1350, 1500, 1550, 1600, 1800, then the median income would be Rs. 1500. Suppose one more person joins and his income is Rs. 1850, then the median income of eight persons would be $\frac{1500+1550}{2}$ = 1525 (since the number of observations is even, the median is the arithmetic mean of the 4th persons and 5th persons).

For grouped data, the following formula may be used to locate the value of median.

$$\text{Med.} = L + \frac{N/2 - pcf}{f} \times i$$

Where L is the lower limit of the median class, pcf is the preceding cumulative frequency to the median class, f is the frequency of the median class and i is the size of the median class.

As an illustration, consider the following data which relate to the age distribution of 1000 workers in an industrial establishment.

Age (Years)	No. of workers	Age (Years)	No. of Workers
Below 25	120	40-45	150
25-30	125	45-50	140
30-35	180	50-55	100
35-40	160	55 and above	25

Determine the median age.

The location of median value is facilitated by the use of a cumulative frequency distribution as shown below in the table 10.4.

Table 10.4: Calculation of Median

Age (Years)	No. of workers f	Cumulative frequency c.f	
Below 25	120	120	
25-30	125	245	
30-35	180	425	
35-40	160	585	Median class
40-45	150	735	
45-50	140	875	
50-55	100	975	
55 and Above	25	1000	
	N = 1000		

Median = size of $\frac{N}{2}$ th observation = $\frac{1000}{2}$ = 500th observation which lies in the class 35 - 40.

Median $= L + \frac{N/2 - pcf}{f} \times i = 35 + \frac{500-425}{160} \times 5$

$= 35 + \frac{375}{160} = 35 + 2.34 = 37.34$ years.

Hence the median age is approximately 37 years. This value of median suggests that half of the workers are below the age of 37 years and other half of the workers are above the age of 37 years.

Advantages of Median

1) The definition of median is rigid which is a quality of a good measure of central tendency.
2) It is easy to understand and calculate.
3) It is not affected by outliers or extreme scores in data.
4) Unless the median falls in an open-ended class, it can be computed for grouped data with open ended classes.
5) In certain cases, it is possible to identify the median through inspection as well as graphically.

Limitations of Median

1) Some statistical procedures using median are quite complex. Computation of the median can be time consuming when large data is involved because the data needs to be arranged in an order before the median is computed.

2) Median cannot be computed exactly when an ungrouped data is even. In such cases, median is estimated as the mean of the scores in the middle of the distribution.

3) It is not based on each and every score in the distribution.

4) It can be affected by sampling fluctuations and thus can be termed as less stable than mean.

10.3.3 Mode

The mode is the typical or commonly observed value in a set of data. It is defined as the value which occurs most often or with the greatest frequency. The dictionary meaning of the term mode is most usual. *For example, in the series of numbers 3, 4, 5, 5, 6, 7, 8, 8, 8, 9, the mode is 8 because it occurs the maximum number of times.*

The calculations are different for the grouped data, where the modal class is defined as the class with the maximum frequency. The following formula is used for calculating the mode.

$$\text{Mode} = L + \frac{d_1}{d_1 + d_2} \times i$$

where L is lower limit of the modal class, d1 is the difference between the frequency of the modal class and the frequency of the preceding class, d2 is the difference between the frequency of the modal class and the frequency of the succeeding class, i is the size of the modal class. To illustrate the computation of mode, let us consider the following data.

Table 10.5 Calculation of Mode

Daily Sales (Rs. thousand)	No. of firms	Daily Sales (Rs. thousand)	No. of firms
20-30	15	60-70	35
30-40	23	70-80	25
40-50	27	80-90	5
50-60	20		

Since the maximum frequency 35 is in the class 60-70, therefore 60-70 is the modal class. Applying the formula, we get

$$\text{Mode} = L + \frac{d_1}{d_1 + d_2} \times i = 60 + \frac{35-20}{(35-20)+(35-25)} \times 10$$

$$= 60 + \frac{150}{25}$$

$$= 60 + 6 = \text{Rs.}66.$$

Hence modal daily sales are Rs. 66.

Advantages of Mode

1) It is not only easy to comprehend and calculate but it can also be determined by mere inspection.
2) It can be used with quantitative as well as qualitative data.
3) It is not affected by outliers or extreme scores.
4) Even if a distribution has one or more than one open ended class(s), mode can easily be computed.

Limitations of Mode

1) It is sometimes possible that the scores in the data vary from each other and in such cases the data may have no mode.
2) Mode cannot be rigidly defined.
3) In case of bimodal, trimodal or multimodal distribution, interpretation and comparison becomes difficult.
4) Mode is not based on the whole distribution.
5) It may not be possible to compute further mathematical procedures based on mode.
6) Sampling fluctuations can have an impact on mode.

10.3.4 Relationship among Mean, Median and Mode

A distribution in which mean, median and mode coincide is known as a symmetrical (bell shaped) distribution. If a distribution is skewed (that is, not symmetrical) then mean, median, and mode are not equal. In a moderately skewed distribution, a very interesting relationship exists among mean, median and mode. In such type of distributions, it can be proved that the distance between mean and median is approximately one third of the distance between the mean and mode. This is shown below in figure 10.1 for two types of such distributions.

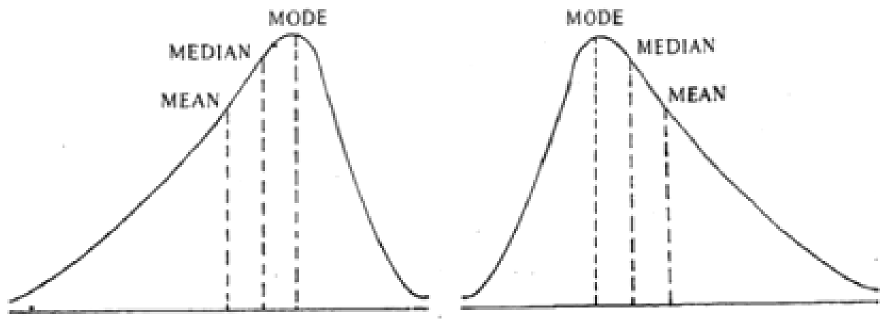


Figure 10.1: Relationship between Mean, Median and Mode

This relationship can be expressed as follows:

Mean - Median = 1/3 (Mean - Mode)

or Mode = 3 Median - 2 Mean

Similarly, we can express the approximate relationship for median in terms of mean and mode. Also, this can be expressed for mean in terms of median and

mode. Thus, if we know any of the two values of the averages, the third value of the average can be determined from this approximate relationship.

For example, consider a moderately skewed distribution in which mean and median is 35.4 and 34.3 respectively. Calculate the value of mode.

To compute the value of mode, we use the approximate relationship

Mode = 3 Median - 2 Mean

= 3 (34.3) - 2 (35.4)

= 102.9-70.8 = 32.1

Therefore, the value of mode is 32.1.

Check Your Progress A

1. Discuss the concept of measures of central tendency with a focus on characteristics of a good measure of central tendency

   ……………………………………………………………………………

   ……………………………………………………………………………

   ……………………………………………………………………………

   ……………………………………………………………………………

   ……………………………………………………………………………

2. Compute mean from the following data:

Class Intervals (Marks)	**Frequencies (*f*)**
50- 59	4
40- 49	5
30- 39	6
20- 29	5
10- 19	5
1- 9	5
	N=30

3. Compute mean, median and mode for the following data:

   23, 34, 43, 65, 67, 67, 78, 65, 43, 34, 45, 33, 23, 67, 60 (N= 15)

10.4 MEASURE OF DISPERSION

In the previous section, measures of central tendency were discussed. While measures of central tendencies are indeed very valuable, their usefulness is rather limited. Although through these measures we can compare the two or more groups, a measure of central tendency is not sufficient for the comparison of two or more groups. They do not show how the individual scores are spread out. Let us take another example, similar to the one that we discussed under the section on introduction. A math teacher is interested to know the performance of two groups (A and B) of his /her students. He/she

gives them a test of 40 points. The marks obtained by the students of groups A and B in the test are as follows:

Marks of Group A: 5,4,38,38,20,36,17,19,18,5 (N = 10, Total = 200, Mean = 20)

Marks of Group B: 22,18,19,21,20,23,17,20,18,22 (N = 10, Total = 200, Mean = 20) The mean scores of both the groups is 20, as far as mean goes there is no difference in the performance of the two groups. But there is a difference in the performance of the two groups in terms of how each individual student varies in marks from that of the other. For instance, the test scores of group A are found to range from 5 to 38 and the test scores of group B range from 18 to 23.

It means that some of the students of group A are doing very well, some are doing very poorly and performance of some of the students is falling at the average level. On the other hand, the performance of all the students of the second group is falling within and near about the average (mean) that is 20. It is evident from this that the measures of central tendency provide us incomplete picture of a set of data. It gives insufficient base for the comparison of two or more sets of scores. Thus, in addition to a measure of central tendency, we need an index of how the scores are scattered around the center of the distribution. In other words, we need a measure of dispersion. It is also known as measure of variability. A measure of central tendency is a summary of scores, and a measure of dispersion is summary of the spread of scores. Information about dispersion or variability is often as important as that about the central tendency.

According to King and Bear (2001), measures of dispersion express quantitatively the extent to which the score in a distribution scatter around or cluster together. They describe the spread of an entire set of scores, they do not specify how far a particular score diverges from the centre of the group. These measures of dispersion do not provide information about the shape of a distribution or the level of performance of a group.

The term variability or dispersion is also known as the average of the second degree, because here we consider the arithmetic mean of the deviations from the mean of the values of the individual items. To describe a distribution adequately, therefore, we usually must provide a measure of central tendency and a measure of variability. Measures of variability are important in statistical inference. With the help of measures of dispersion, we can know about fluctuation in random sampling. How much fluctuation will occur in random sampling? This question in fundamental to every problem in statistical inference, it is a question about variability.

The measures of variability are important for the following purposes:

- Measures of variability are used to test the extent to which an average represents the characteristics of a data. If the variation is small then it indicates high uniformity of values in the distribution and the average represents the characteristics of the data. On the other hand, if variation is large then it indicates lower degree of uniformity and unreliable average.

- Measures of variability help in identifying the nature and cause of variation. Such information can be useful to control the variation.
- Measures of variability help in the comparison of the spread in two or more sets of data with respect to their uniformity or consistency.
- Measures of variability facilitate the use of other statistical techniques such as correlation, regression analysis, and so on.

Functions of Variability:

The major functions of dispersion or variability are as follows:

- It is used for calculating other statistics such as analysis of variance, degree of correlation, regression etc.
- It is also used for comparing the variability in the data obtained as in the case of Socio-Economic Status, income, education etc.
- To find out if the average or the mean/median/mode worked out is reliable. If the variation is small then we could state that the average calculated is reliable, but if variation is too large, then the average may be erroneous.
- Dispersion gives us an idea if the variability is adversely affecting the data and thus helps in controlling the variability.

The measures of dispersion most commonly used in statistics are as follow:

1) Range
2) Mean Deviation
3) Standard Deviation
4) Coefficient of Variation

10.4.1 Range

Range can be defined as the difference between the highest and lowest score in the distribution. This is calculated by subtracting the lowest score from the highest score in the distribution. The equation is as follows:

Range = Highest Score – Lowest Score(R=H-L)

The range is a rough measure of dispersion because it tells about the spread of the extreme scores and not the spread of any of the scores in between. For instance, the range for the distribution 4,10,12,20, 25, 50 will be 50 - 4 = 46.

Let us understand the steps in computation of range with the help of an example,

For example, if there are 10 students who have obtained marks in history as mentioned below:

50, 45, 42, 46, 55, 54, 59, 60, 62, 64

Step 1: Arrange the scores in ascending order.
42, 45, 46, 50, 54, 55, 59, 60, 62, 64

Step 2: Identify the lowest and the highest score in the data In the above data, the lowest score is 42 and the highest score is 64.

Step 3: Compute range with the help of the following formula:

R = H - L

64 – 42 = 22.
Thus, the range obtained is 22.

Advantages of Range:

1) It is easiest to compute when compared with other measures of variability and its meaning is direct.
2) The range is ideal for preliminary work or in other circumstances where precision is not an important requirement.
3) It is quite useful in case where the purpose is only to find out the extent of extreme variation, such as temperature, rainfall etc.
4) Range is effectively used in the application of tests of significance with small samples.

Limitations of Range:

1) The calculation of range is based only on two extreme values in the data set and does not consider other values of the data set. Sometimes, the extreme values of the two different data sets may be same or similar, but the two data sets may be differ in dispersion.
2) Its value is sensitive to change in sampling. The range varies more with sampling fluctuation. That is different sample of the same size from the same population may have different range.
3) Its value is influenced by large samples. In many types of distribution, including normal distribution, the range is dependent on sample size. The sampling variance increases rapidly with increase in sample size.
4) Range cannot be used for open-ended class intervals since the highest and the lowest scores of the distribution are not available and thus the range cannot be computed.
5) Further mathematical calculations are not possible for range.
6) Range indicates two extreme scores, thus the magnitude or frequency of intermediate scores is missing.
7) It does not indicate the form of distribution, like skewness, kurtosis, or modal distribution of scores.
8) A single extreme score may also increase the range disproportionately.

Uses of the Range

Range is applied in diverse areas discussed as follows:

1) Range is used in areas where there are small fluctuations, such as stock market, rate of exchange, etc.

2) Range may be used in day-to-day activities like, daily sales in a grocery store, monthly wages in a factory, etc.
3) Range is used in weather forecasts, like variation in temperature in a day.
4) When the researcher is only interested in the extreme scores or total spread of the scores, range is the most useful measure of variability.
5) Range can also be used when the data are too scant or too scattered to justify the use of most appropriate measure of variability.

10.4.2 Mean Deviation

Mean deviation is the average of difference of the values of items from some average of the series. In other words, it is assessed on the basis of the average of the distribution. Such a difference is technically described as deviation. It is utilized in checking the spread of data with respect to the central value. In calculating mean deviation, we ignore the minus sign of deviations while taking their total for obtaining the mean deviation. Mean deviation is calculated as follows:

Example: Suppose we have a set of observations given by {2, 7, 5, 10} and we want to calculate the mean deviation about the mean. We find the mean of the data given by 6. Then we subtract the mean from each value, take the absolute value of each result and add them up to get 10. Finally, we divide this value by the total number of observations (4) to get the mean deviation as 2.5.

Table 10.6: Mean Deviation Formulas

	Ungrouped Data	**Grouped Data**
About Mean	$\frac{\Sigma \lvert Xi - Mean \rvert}{n}$	$\frac{\Sigma fi \lvert Xi - Mean \rvert}{\Sigma fi}$
About Median	$\frac{\Sigma \lvert Xi - Median \rvert}{n}$	$\frac{\Sigma fi \lvert Xi - Median \rvert}{\Sigma fi}$

Depending upon the type of data available as well as the type of the central point, there can be several different formulas to calculate the mean deviation. Given below are the different mean deviation formulas.

Mean Deviation Formula for Ungrouped Data

Data that is not sorted or classified into groups and remains in raw form is known as ungrouped data. To calculate the mean deviation for ungrouped data the formula is as follows:

$$\text{MAD} = \frac{\Sigma \lvert xi - \bar{x} \rvert}{n}$$

Here, xi represents the i^{th} observation, $\bar{x}$ represents the central point (mean, median, or mode), and 'n' is the number of observations in the data set.

Mean Deviation Formula for Grouped Data

When data is organized and classified into groups it is known as grouped data. Grouping of data is done by continuous and discrete frequency distributions. The mean deviation formulas for grouped data are given below:

Mean Deviation for Continuous Frequency Distribution:

Such a type of grouped data consists of class intervals. The frequency of repetition of an observation within each interval is given by the continuous frequency distribution. The mean deviation formula is as follows:

$$\text{MAD} = \frac{\sum fi|xi - \bar{x}|}{\sum fi}$$

Here, fi is the frequency of repetition of xi. xi denotes the mid value of the class interval.

Mean Deviation for Discrete Frequency Distribution:

In this type of data, the individual data points are specified and the frequency with which they occur is also mentioned. To calculate the mean deviation for a discrete frequency distribution, the formula is given as follows:

$$\text{MAD} = \frac{\sum fi|xi - \bar{x}|}{\sum fi}$$

Here, xi denotes the specified individual value and fi is the frequency of occurrence of that value.

Mean Deviation about Mean:

The mean is also known as the expected value of a data set. The simple definition of mean is given as the sum of all observations divided by the total number of observations. The formulas for mean deviation about the mean are given below:

- Ungrouped data MAD = $\frac{\sum|xi-\mu|}{n}$

  where, mean is μ = $\frac{X1+X2+\ldots+Xn}{n}$

- Continuous and discrete frequency distribution MAD = $\frac{\sum fi|xi-\mu|}{\sum fi}$

  where, mean of grouped data is μ = $\frac{\sum xifi}{\sum fi}$

Mean deviation about median:

- Ungrouped data MAD = $\sum\frac{|\text{xi}-\text{M}|}{\text{n}}$

  where, if n is odd, then median M = $\frac{(\text{n}+1)^{\text{th}}}{2}$ observation.

  if n is even, then median M = $\frac{\frac{\text{n}}{2}\text{th}_{\text{obs}} + \left(\frac{\text{n}}{2}+1\right)\text{th}_{\text{obs}}}{2}$

- Discrete frequency distribution MAD = $\frac{\sum\text{fi}|\text{xi}-\text{M}|}{\sum\text{fi}}$

The median is calculated in the same way as ungrouped data.

- Continuous frequency distribution MAD = $\frac{\sum fi|xi-M|}{\sum fi}$

Median of grouped data M $= l + \frac{\sum fi}{2} - \frac{cf}{f} \times h$

cf stands for cumulative frequency preceding the median class, l is the lower value of the median class, h is the length of the median class and f is the frequency of the median class.

Mean Deviation about Mode:

Mode is defined as that value that occurs most frequently in a given data set. The formulas to calculate mean deviation about mode are as follows:

- Ungrouped data MAD $\frac{= \sum|xi-mode|}{n}$

  where mode = the most frequently occurring value in a data set.

- Discrete frequency distribution MAD = $\frac{\sum fi|xi-mode|}{\sum fi}$

The mode can be calculated in the same way as ungrouped data

- Continuous frequency distribution MAD $= \frac{\sum fi|xi-mode|}{\sum fi}$

where, mode of grouped data = $l + \left(\frac{f-f1}{2f-f1-f2}\right) \times h$

l is the lower value of the modal class, h is the size of the modal class, f is the frequency of the modal class, f1 is the frequency of the class preceding the modal class, and f2 is the frequency of the class succeeding the modal class.

Regardless, of whether the mean deviation about the mean, median or mode needs to be determined, the general steps remain the same. The only difference will be in the formulas used to calculate the mean, median or mode depending upon the type of data available to us.

Example: Suppose the mean deviation about the mean has to be determined for the data set {10, 15, 17, 15, 18, 21}. Then the below-given steps can be followed.

- **Step 1:** Calculate the value of the mean, mode, or median of the given data values. Here, we find the mean given by 16.
- **Step 2:** Subtract the value of the central point (here, mean) from each data point. (10 - 16), (15 - 16), ..., (21 - 16) = -6, -1, 1, -1, 2, 5.
- **Step 3:** Now take the absolute of the values obtained in step 2. The values are 6, 1, 1, 1, 2, 5
- **Step 4:** Take the sum of all the values obtained in step 3. This gives 6 + 1 + 1 + 1 + 2 + 5 = 16
- **Step 5:** Divide this value by the total number of observations. This results in the mean deviation. As there are 6 observations hence, 16 / 6 = 2.67 which is the mean deviation about the mean.

Advantages of Mean Deviation:

1) It is easy to calculate and simple to understand.
2) It does not get extremely affected by outliers.
3) It is widely used in business and commerce.
4) It has the least sample fluctuations as compared to other statistical measures.
5) It is a good comparison measure as it is based on the deviations from the mid-value.

Limitations of Mean Deviation:

1) It is not capable of further algebraic treatment hence, this can lead to reduced usability.
2) It is not rigidly defined as it can be calculated with respect to mean, median, and mode.
3) Negative and positive signs are ignored because we take the absolute value. This can lead to inaccuracies in the result.

10.4.3 Standard Deviation

The term standard deviation (SD) was first used in writing by Karl Pearson in 1894. The standard deviation of population is denoted by 'σ' (Greek letter sigma) and that for a sample is 's'**.** A useful property of SD is that unlike variance it is expressed in the same unit as the data. This is most widely used method of variability. The standard deviation indicates the average of distance of all the scores around the mean. It is the positive square root of the mean of squared deviations of all the scores from the mean. It is also called positive square root of variance, and called as 'root mean square deviation'. Mangal (2002, page 71) defined standard deviation as "as the square root of the average of the squares of the deviations of each score from the mean". SD is an absolute measure of dispersion and it is the most stable and reliable measure of variability.

Standard deviation shows how much variation there is, from the mean. SD is calculated from the mean only. If standard deviation is low it means that the data is close to the mean. A high standard deviation indicates that the data is spread out over a large range of values. Standard deviation may serve as a measure of uncertainty. If you want to test the theory or in other word, want to decide whether measurements agree with a theoretical prediction, the standard deviation provides the information. If the difference between mean and standard deviation is very large then the theory being tested probably needs to be revised. The mean with smaller standard deviation is more reliable than mean with large standard deviation. A smaller SD shows the homogeneity of the data. The value of standard deviation is based on every observation in a set of data. It is the only measure of dispersion capable of algebraic treatment therefore, SD is used in further statistical analysis. Standard deviation can be computed for ungrouped and grouped data.

Standard deviation for ungrouped data can be computed by the following formula:

SD =$\sqrt{\sum x^2/N}$

The above formula can be explained by the following example.

Let us understand the steps in computation of standard deviation for ungrouped data with the help of example given below in table 10.7.

Table 10.7: Calculation of Standard Deviation

Scores (X)	Deviation from the mean (x)	Deviation square (x^2)
52	-8	64
50	-10	100
56	-4	16
68	8	64
65	5	25
62	2	4
57	-3	9
70	10	100
Total=480		$\sum x^2$= 382
Mean=60		

Step 1: Add all the scores ($\sum x$) and divide this sum by the number of scores (N) and find out mean. Mean of the given scores is M=$\sum X/N$ = 480/8 =60.

Step 2: Find out deviation x by computing X – x, as given in second column above.

Step 3: Square all the deviation to get x^2.

Step 4: Add all the squared deviation to get $\sum x^2$.

Step 5: Compute standard deviation with the help of the formula: SD= $\sqrt{(\sum x^2)}$ /N

= $\sqrt{}$ 382/8 =$\sqrt{}$ 47.7 =6.91

Thus, the standard deviation for this data is 6.91.

Computations of SD for Grouped Data by Long Method

Standard deviation of grouped data can be computed by the formula,

SD= $\sqrt{\sum f\mathbf{x}^2}$**/N** Where,

$\sum fx^2$= When frequencies (f) are multiplied with their respective deviation squared (x^2), fx^2 is obtained. Total of all fx^2 is $\sum fx^2$.

N= Total number of scores

Let us understand the steps in computation of standard deviation for grouped data with the help of example given below in table 10.8.

Table 10.8: Calculation of Standard Deviation of Grouped Data

Class interval	**Frequency (*f*)**	**Midpoint (X)**	**fx**	**Deviation of midpoint from the mean (x= X- M)**	**Deviation squared (x^2)**	***f*x2**
(1)	(2)	(3)	(4)	(5)	(6)	(7)
127-129	1	128	128	17.6	309.76	309.76
124-126	2	125	250	14.6	213.16	426.32
121-124	2	122	244	11.6	134.56	269.12
118-120	2	119	238	8.6	73.96	147.92
115-117	4	116	464	5.6	31.36	125.44
112-114	4	113	452	2.6	6.76	27.04
109-111	4	110	440	-0.4	0.16	0.64
106-108	2	107	214	-3.4	11.56	23.12
103-106	2	104	208	-6.4	40.96	81.92
100-102	2	101	202	-9.4	88.36	176.72
	Total=25		**∑fx=2760**			**∑*f*x2=1588**

Step 1: Midpoint is computed for respective class intervals and entered in column 3 as shown in the above table.

Step 2: *f*X is then computed by multiplying the frequencies and the midpoint, the values thus obtained are entered in column 4.

Step 3: Add all the scores under *f*X and divide this sum by the number of scores (N) and find out mean. Thus, M= ∑*f*X/ N, that is, 2760/ 25= 110.4. The mean obtained is 110.4.

Step 4: x is now computed by subtracting the M from X (midpoint). The values thus obtained are entered in column 5.

Step 5: The x is then squared to obtain x2 (column 6).

Step 6: *f*x2 is then computed by multiplying f and x2. The values thus obtained are entered in column 7.

Step 7: Add the fx^2 to obtain ∑fx2. In the present example it is obtained as 1588.

Step 8: Compute standard deviation with help of the formula: **SD=** $\sqrt{\sum fx^2/N}$

$= \sqrt{1588/25} = \sqrt{63.52} = 7.97$

The standard deviation thus obtained is 7.97.

Standard deviation from grouped data can also be computed by the following formula.

SD= i $\sqrt{\sum fx'^2/N - (\sum(fx')/N)^2}$

where,

i= The size of the class interval

$\sum fx'^2$=When frequencies (f) are multiplied with their respective deviation squared (x'^2), fx'^2 is obtained. Total of all fx'^2 is $\sum fx'^2$.

Let us understand the steps in computation of standard deviation for grouped data by short method with the help of example given below in table 10.9:

Table 10.9: Standard Deviation for Grouped Data by Short Method

Class interval	**Frequency (*f*)**	**Midpoint (X)**	**Deviation of midpoint from the mean (x= X- AM)/i**	**Deviation squared (fx')**	***f*x2**
(1)	(2)	(3)	(5)	(6)	(7)
127-129	1	128	4	4	309.76
124-126	2	125	3	6	426.32
121-124	2	122	2	4	269.12
118-120	2	119	1	2	147.92
115-117	4	116	0	0	125.44
112-114	4	113	-1	-4	27.04
109-111	4	110	-2	-8	0.64
106-108	2	107	-3	-6	23.12
103-106	2	104	-4	-8	81.92
100-102	2	101	-5	-10	176.72
	Total=25			$\sum fx'$=-20	$\sum fx'^2$ =352

Step 1: Find out midpoint of each class interval, that is entered in column 3.

Step 2: Assume one value as mean. In this example, assumed mean is taken as 116.

Step 3: Find out the difference between midpoint and assumed mean and divide it by class intervals to get x′, (x' = X- AM/i) and enter the obtained value in column 5.

Step 4: Multiply each x′ by respective frequency and get fx' (column 6).

Step 5: fx′ is squared to get fx'^2. Add all the fx'^2 to obtain $\sum fx'$.

Step 6: Compute standard deviation with the help of the formula

SD= i √ $\sum fx'^2$/N-($\sum$(fx′)/N)^2

= 3 √352/ 25- (-20/ 25)2 = 3 √14.08 - (-0.8)2
= 3 √ 14.08 - 0.64
= 3 √3.44

= 3 x 1.85 = 5.55
The standard deviation thus obtained is 5.55.

Merits of Standard Deviation

1) It is widely used because it is the best measure of variation by virtue of its mathematical characteristics.
2) It is based on all the observations of the data.
3) It gives an accurate estimate of population parameter when compared with other measures of variation.
4) SD is least affected by sample fluctuations
5) It is also possible to calculate combined SD, that is not possible with other measures.
6) Further statistics can be applied on the basis of SD like, correlation, regression, tests of significance, etc.
7) Coefficient of variation is based on mean and SD. It is the most appropriate method to compare variability of two or more distributions.

The limitations of SD are as follows:

1) While calculating standard deviation more weight is given to extreme values and less to those, near the mean.
2) It is difficult to compute as compared to other measures of dispersion.

Uses of Standard Deviation

1) SD is used when one requires a more reliable and accurate measure of variability but it is recommended when the distribution is normal or near to normal.
2) It is used when further statistics like, correlation, regression, tests of significance, etc. have to be computed.

10.4.4 Coefficient of Variation

The relative measure corresponding to SD is the coefficient of variation. It is a relative measure of dispersion developed by Karl Pearson. When we want to compare the variations (dispersion) of two different series, relative measures of standard deviation must be calculated. This is known as co-efficient of variation or the co-efficient of SD. It is defined as the SD expressed as a percentage of the mean.The coefficient of variation represents the ratio of the standard deviation to the mean, and it is a useful statistic for comparing the degree of variation from one data series to another, even if the means are drastically different from each other. Thus, it is more suitable than SD or variance. It is given as a percentage and is used to compare the consistency or variability of two or more data series.

The formula for computing coefficient of variation is as follows:

$V = 100 \times \sigma / M$

Where,
V = Variance
σ = Standard deviation

M = Mean

To understand the computation with the help of an example,

If the standard deviation of marks obtained by 10 students in a class test in English is 10 and Mean is 79, then,

V= 100 × 10 / 79 = 1000 / 79
= 12.65

Check Your Progress B

1) What is the formula for coefficient of Variation?

....

....

....

....

....

2) List the steps in computation of standard deviation for ungrouped data.

....

....

....

....

3) What are the advantages of Standard Deviation ?

....

....

....

....

....

4) Explain the Uses of Range?

....

....

....

....

....

....

10.5 LET US SUM UP

Descriptive statistics are essential tools that help us make sense of large amounts of data by summarizing and simplifying it into meaningful insights. They allow us to spot patterns, trends, and relationships, making it easier to understand raw information and use it for decision-making.

At the heart of descriptive statistics are measures of central tendency: the mean, median, and mode. These provide a single value that captures the essence of an entire dataset. The mean, or average, gives us a general idea of the data's center, the median pinpoints the middle value when data is sorted, and the mode shows the most frequently occurring value. Each of these measures has its strengths and is best suited for specific kinds of data.

Alongside these, measures of dispersion describe how spread out or variable the data is. For example, the range calculates the difference between the highest and lowest values, while the standard deviation tells us how much the data deviates, on average, from the mean. The coefficient of variation helps compare the relative variability of different datasets by expressing the standard deviation as a percentage of the mean.

These statistical tools are powerful for comparing groups, checking data quality, and guiding decisions based on evidence. Whether it's understanding student performance in education, analyzing customer behavior in marketing, or optimizing operations in business, descriptive statistics make complex data easy to interpret. Visual tools like charts and graphs further enhance communication, making findings accessible to everyone.

While they focus on summarizing data, descriptive statistics pave the way for deeper analyses and predictions. Their simplicity and versatility make them invaluable in areas like business, healthcare, education, and research, helping professionals make informed decisions and uncover hidden insights.

10.6 KEYWORDS

Arithmetic Mean: The average of a dataset, calculated as the sum of values divided by the number of values.

Coefficient of Variation: A relative measure of variability, showing standard deviation as a percentage of the mean.

Data Dispersion: The extent to which data values are spread out around a central point.

Distribution: The pattern of data points across a dataset.

Frequency: The number of times a value occurs in a dataset.

Grouped Data: Data organized into intervals or categories.

Median: The middle value of a dataset when arranged in ascending order.

Mode: The value that appears most frequently in a dataset.

Outlier: A data point significantly different from others in a dataset.

Range: The difference between the highest and lowest values in a dataset.

Standard Deviation: A measure of the average distance of data points from the mean.

Ungrouped Data: Data presented as individual points without categorization.

Variability: The extent to which data points differ from each other.

10.7 ANSWER TO CHECK YOUR PROGRESS

Check Your Progress A:

2 Answer. 28.92

3 Answer Mean 49.8, Median 45 Mode 67

10.8 TERMINAL QUESTIONS

1. Explain the significance of descriptive statistics in data analysis with relevant examples. Discuss its role in summarizing and simplifying large datasets.
2. Discuss the concept of measures of central tendency with a focus on characteristics of a good measure of central tendency.
3. Explain the concept of the arithmetic mean. How is it calculated for grouped data? Provide examples.
4. What is the range? Discuss its advantages, limitations, and applications in measuring data variability
5. Compare and contrast measures of central tendency and measures of dispersion. Discuss how both complement each other in understanding datasets comprehensively.
6. Compute the range, and standard deviation from the following ungrouped data:

   a) 30, 35, 36, 39, 42, 46, 38, 34, 35

   b) 52, 50, 56, 68, 65, 62, 57, 70

   Answer: a) Range: 16, Standard Deviation:4.44

   b) Range 20, Standard Deviation 6.91

10.9 FURTHER READINGS

1. Nick, T. G. (2007). Descriptive statistics. *Topics in biostatistics*, 33-52.
2. Kaur, P., Stoltzfus, J., & Yellapu, V. (2018). Descriptive statistics. *International Journal of Academic Medicine*, *4*(1), 60-63.
3. Fisher, M. J., & Marshall, A. P. (2009). Understanding descriptive statistics. *Australian critical care*, *22*(2), 93-97.
4. Holcomb, Z. (2016). *Fundamentals of descriptive statistics*. Routledge.
5. Pérez-Vicente, S., & Ruiz, M. E. (2009). Descriptive statistics. *Allergologia et immunopathologia*, *37*(6), 314-320.

UNIT 11 FORMULATION AND TESTING OF HYPOTHESES

Structure

11.0 OBJECTIVES

After studying this unit, you will be able to:

- Describe hypothesis, its formulation and significance;
- Understand and explain hypothesis testing steps;
- Explain the types of parametric and non-parametric tests; and
- Understand the concepts of correlation and regression.

11.1 INTRODUCTION

In the previous unit, we explored how descriptive statistics helps in summarized, organized data using various statistical measures such as mean, median and standard deviation. Descriptive statistics help researchers understand the patterns and distribution within data set, providing valuable insight into characteristics. However, while descriptive statistics allow us to describe and visualize dates, they do not provide conclusive evidence about relationships between variables or allow us to make prediction. This is where hypothesis formulation and testing come into play. Formulation and testing

of a hypothesis are fundamental steps in scientific research and statistical analysis. The process begins with the formulation of a hypothesis, which is a clear and testable statement or prediction about the relationship between two or more variables. This hypothesis is typically based on existing theories, observations, or previous research, and it provides a direction for investigation.

There are two main types of hypotheses: the null hypothesis (H0) and the alternative hypothesis (H1 or Ha). The null hypothesis posits that there is no significant effect or relationship between the variables, serving as a default assumption. In contrast, the alternative hypothesis suggests that there is a significant effect or relationship, which the research aims to support.

Once the hypotheses are formulated, the next step is hypothesis testing. This involves collecting data and using statistical methods to determine whether the evidence supports rejecting the null hypothesis in favour of the alternative hypothesis. The process includes selecting an appropriate statistical test, determining the significance level (often denoted as alpha, typically set at 0.05), and calculating the p-value, which indicates the probability of observing the data if the null hypothesis is true.

If the p-value is less than the significance level, the null hypothesis is rejected, suggesting that the data provides enough evidence to support the alternative hypothesis. Conversely, if the p-value is greater than the significance level, the null hypothesis is not rejected, implying that the data does not provide sufficient evidence to support the alternative hypothesis.

The formulation and testing of a hypothesis are critical for drawing valid conclusions from research, allowing scientists and researchers to systematically evaluate their predictions and contribute to the body of knowledge in their field.

11.2 SIGNIFICANCE OF HYPOTHESES FORMULATION

Hypotheses is usually considered the principal instrument in research. Its main function is to suggest new experiments and observations. In fact, many experiments are carried out with the deliberate object of testing hypotheses. Decision-makers often face situations wherein they are interested in testing hypotheses on the basis of available information and then take decisions on the basis of such testing. In social science, where direct knowledge of population parameter(s) is rare, hypothesis testing is the often-used strategy for deciding whether a sample data offer such support for a hypothesis that generalization can be made. Thus, hypothesis testing enables us to make probability statements about population parameter(s). The hypothesis may not be proved absolutely, but in practice it is accepted if it has withstood critical testing. Before we explain how hypotheses are tested through different tests meant for the purpose, it will be appropriate to explain clearly the meaning of a hypothesis and the related concepts for better understanding of the hypothesis testing techniques.

11.3 FORMULATION OF HYPOTHESES

Formulating a hypothesis is a systematic process that involves several key steps to ensure that the hypothesis is clear, testable, and relevant to the research question.

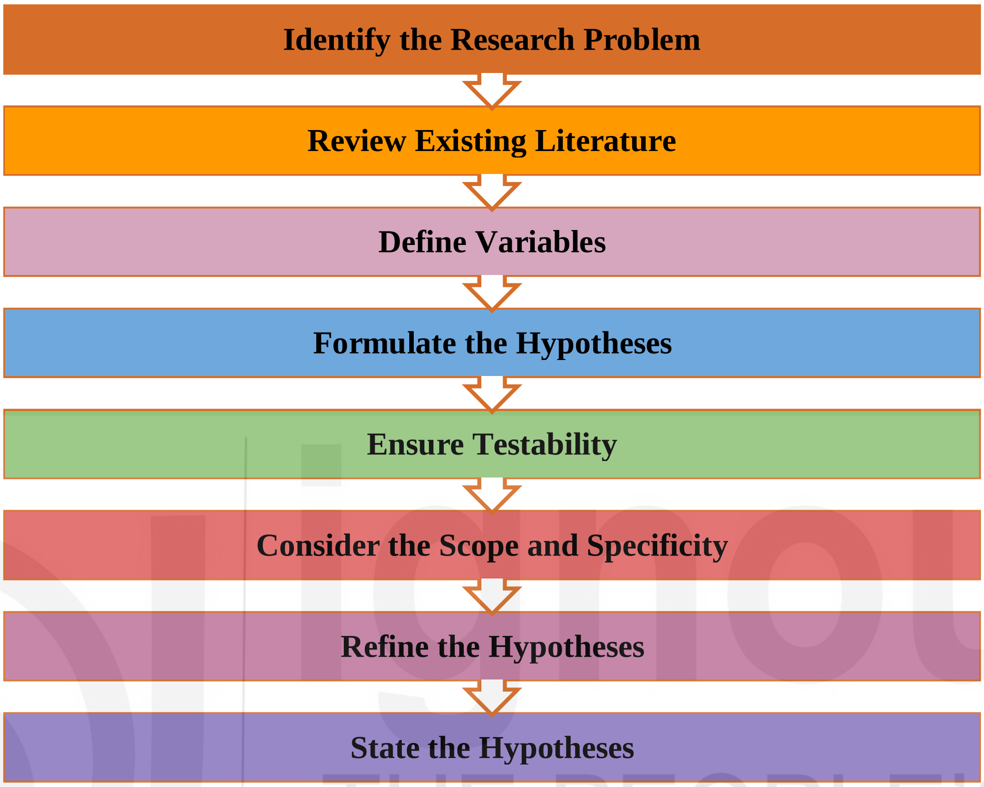


Figure 11.1: Steps of Hypotheses Formulation

Here are the steps typically followed in the formulation of a hypothesis:

1. **Identify the Research Problem:** Hypotheses Formulation begins with clearly defining the research problem or question. This involves understanding the broader topic of interest and narrowing it down to a specific issue or question that you want to explore.

2. **Review Existing Literature:** After identifying the research problem, a thorough review of the existing literature on the topic is conducted in this step. This helps you understand what is already known about the subject, identify gaps in the current knowledge, and provide a theoretical basis for your hypothesis.

3. **Define variables:** In this step, identify key variables involved in your research. Determine which variables are independent (those you will manipulate or consider as the cause) and which are dependent (those you will measure or consider as the effect).

4. **Formulate the Hypotheses:** Develop a clear and concise statement that predicts the relationship between the independent and dependent variables. This statement should be specific and testable, meaning it can be supported or refuted through empirical evidence.

- **Null hypothesis (H0):** Formulate the null hypothesis, which asserts that there is no effect or relationship between the variables.

- **Alternative hypothesis (H1 or Ha):** Formulate the alternative hypothesis, which suggests that there is an effect or relationship between the variables.

5. **Ensure Testability:** Make sure that the hypothesis is testable. This means that it should be possible to gather data that either supports or refutes the hypothesis through observation, experimentation, or other research methods.
6. **Consider the scope and specificity:** After ensuring the testability of the hypothesis, you should make sure that the hypothesis is neither too broad nor too narrow. A well-defined hypothesis should be specific enough to be tested within the scope of your research but broad enough to be meaningful.
7. **Refine the Hypotheses:** Review and refine the hypothesis to ensure clarity and precision. It should be free of ambiguity, clearly stating the expected relationship between the variables.
8. **State the Hypotheses:** This step requires expressing the hypothesis in a formal statement, and this statement should be included in your research proposal or study design, providing a foundation for your research.

By following these steps, researchers can formulate a hypothesis that is well-defined, testable, and aligned with the goals of their study, setting the stage for meaningful and scientifically valid research.

	Decision	
	Accept H_0	**Reject H_0**
H_0 (true)	**Correct decision**	**Type I error (α error)**
H_0 (false)	**Type II error (β error)**	**Correct decision**

11.4 STEPS OF HYPOTHESES TESTING

Hypotheses testing is a critical process in statistics that allows researchers to make inferences about a population based on sample data. The process involves several key steps to determine whether the observed data provides enough evidence to reject the null hypothesis in favor of the alternative hypothesis.

Here are the steps involved in hypothesis testing:

1. **State the hypotheses**:

- **Null hypothesis (H0)**: This is the default assumption that there is no effect or relationship between variables. It serves as the baseline hypothesis.

- **Alternative hypothesis (H1 or Ha)**: This is the hypothesis that there is an effect or relationship between the variables. The goal of hypothesis testing is to determine whether there is enough evidence to reject the null hypothesis in favor of the alternative.

2. **Choose the significance level (α)**:

- The significance level, typically denoted as alpha (α), is the probability of rejecting the null hypothesis when it is actually true (Type I error). Common choices for α are 0.05, 0.01, or 0.10, with 0.05 being the most widely used.

FLOW CHART FOR HYPOTHESIS TESTING

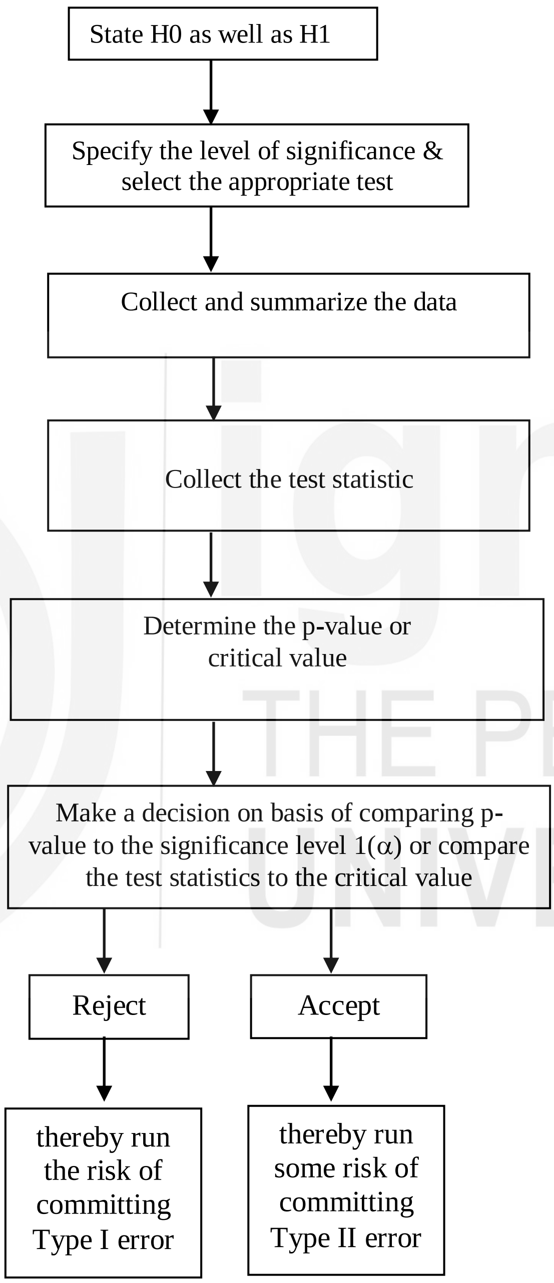


Figure 11. 2: Flow Chart of Hypotheses Testing

Select the appropriate test:

- Choose the appropriate statistical test based on the data type, data distribution, and research question. Examples of common tests include the t-test, z-test, chi-square test, and ANOVA. The choice of test depends on factors such as whether the data is categorical or continuous and whether you are comparing means, proportions, or other statistics.

3. **Collect and summarize the data**:

   - Gather the sample data through experimentation, observation, or survey. Then, summarize the data using descriptive statistics (e.g., mean, standard deviation) and prepare it for analysis.

4. **Calculate the test statistic**:

   - Use the selected statistical test to calculate the test statistic, which measures the degree to which the sample data is consistent with the null hypothesis. The formula for the test statistic varies depending on the type of test being used.

5. **Determine the P-value or critical value**:

   - **P-Value**: Calculate the p-value, which represents the probability of obtaining a test statistic at least as extreme as the one observed, assuming the null hypothesis is true.

   - **Critical Value**: Alternatively, you can determine the critical value(s) corresponding to the chosen significance level. The critical value defines the cutoff points beyond which the null hypothesis will be rejected.

6. **Make a Decision**:

   - Compare the p-value to the significance level (α):

     - If the p-value is less than or equal to α, reject the null hypothesis (H0). This suggests that there is enough evidence to support the alternative hypothesis (H1).

     - If the p-value is greater than α, do not reject the null hypothesis. This suggests that there is not enough evidence to support the alternative hypothesis.

   - Alternatively, compare the test statistic to the critical value(s):

     - If the test statistic falls in the rejection region (beyond the critical value), reject the null hypothesis.

     - If the test statistic does not fall in the rejection region, do not reject the null hypothesis.

7. **Draw Conclusions**:

   - Based on the decision, draw conclusions about the research question. If the null hypothesis is rejected, conclude that there is evidence to support the alternative hypothesis. If the null hypothesis is not rejected, conclude that there is not enough evidence to support the alternative hypothesis.

8. **Report the Results**:

   - Clearly report the findings, including the test statistic, p-value, and whether the null hypothesis was rejected or not. Provide context for the results, explaining what they mean in terms of the original research question.

By following these steps, researchers can systematically evaluate their hypotheses and draw valid conclusions based on the data, contributing to the body of knowledge in their field.

11.5 PARAMETRIC TEST

Parametric tests are statistical tests that assume the underlying data follows a certain distribution, typically a normal distribution, and rely on parameters such as the mean and standard deviation to make inferences about the population. These tests are powerful and efficient when the assumptions are met because they use more information from the data.

Key Assumptions of Parametric Tests:

1. **Normality**: The data should be approximately normally distributed, especially when the sample size is small.
2. **Homogeneity of Variance**: The variance within each group being compared should be roughly equal.
3. **Independence**: Observations should be independent of each other.
4. **Interval or Ratio Level**: The data should be measured on an interval or ratio scale, where meaningful arithmetic operations can be performed.

Types of Parametric Tests:

1. **T-Tests**:
   - **One-Sample T-Test**: One sample T-Test compares the mean of a single sample to a known value (e.g., population mean). It tests whether the sample mean is significantly different from the population mean.
   - **Independent Samples T-Test (Two-Sample T-Test)**: One sample T-Test compares the means of two independent groups to determine if there is a significant difference between them. It is commonly used when comparing the means of two different groups.
   - **Paired Samples T-Test (Dependent T-Test)**: Paired samples T-Test compares the means of two related groups, such as measurements taken from the same subjects at two different times or under two different conditions.

   We will study about the T-test in detail in the coming head 11.5.2.

2. **ANOVA (Analysis of Variance)**:
   - **One-Way ANOVA**: It tests whether there are significant differences between the means of three or more independent groups based on one factor or independent variable. We will study about one-way ANOVA test in detail in the coming head 11.5.3.
   - **Two-Way ANOVA**: Tests the effect of two independent variables on a dependent variable and also examines the interaction between the two independent variables.
   - **Repeated Measures ANOVA**: Similar to the paired t-test but used when comparing the means of three or more related groups or conditions.

3. **Z-Test**:

- **One-Sample Z-Test**: Similar to the one-sample t-Test, but used when the sample size is large (typically n > 30) and the population variance is known. It compares the sample mean to a known population mean.
- **Two-Sample Z-Test**: Compares the means of two independent samples when the population variances are known.

We will study about the z-test in detail in the coming head 11.5.1.

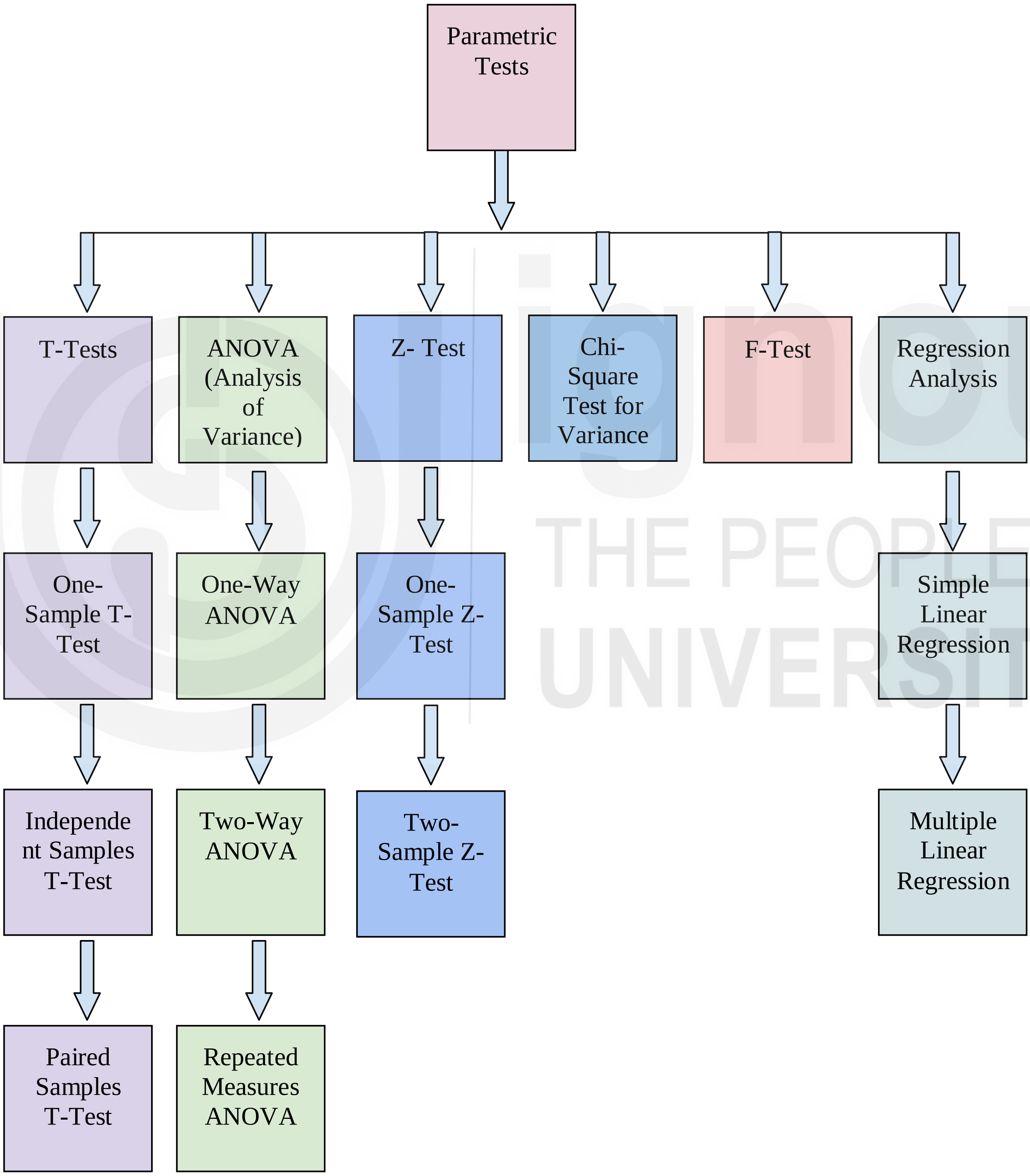


Figure 11.3: Types of Parametric Tests

4. **Chi-Square Test for Variance**:
   - Tests whether the variance of a sample is significantly different from a known or hypothesized population variance. It is used when the data is normally distributed and the sample variance is being compared to a theoretical variance.
5. **F-Test**:
   - Compares the variances of two independent samples to determine if they are significantly different. It is often used in the context of ANOVA and in comparing the fit of two models.
6. **Regression Analysis**:
   - **Simple Linear Regression**: Examines the relationship between two continuous variables, predicting the value of the dependent variable based on the independent variable.
   - **Multiple Linear Regression**: Extends simple linear regression by examining the relationship between one dependent variable and multiple independent variables.

Applications of Parametric Tests:

Parametric tests are widely used in various fields such as psychology, biology, medicine, economics, and social sciences. They are particularly useful when the assumptions about the data distribution are met, as they tend to be more powerful than non-parametric tests, meaning they are more likely to detect a true effect when one exists.

Advantages of Parametric Tests:

- **Efficiency**: They make full use of the data by considering means and variances, leading to more precise estimates.
- **Power**: Parametric tests are generally more powerful than non-parametric tests, meaning they have a higher probability of detecting a true effect when one exists.
- **Wide Applicability**: They are well-suited for many common types of data and research designs.

Limitations of Parametric Tests:

- **Assumption Sensitivity**: They require strict adherence to their assumptions, and violations of these assumptions can lead to incorrect conclusions.
- **Not Suitable for All Data Types**: They are not appropriate for ordinal data or data that does not meet the normality assumption.

In summary, parametric tests are a fundamental part of statistical analysis, providing powerful tools for hypothesis testing when their underlying assumptions are met. They allow researchers to make inferences about population parameters, such as means and variances, based on sample data.

11.5.1 Z Test

The Z-test is a parametric statistical test used to determine whether there is a significant difference between sample and population means or between two sample means, especially when the sample size is large (typically n > 30) and the population variance is known. The Z-test assumes that the data follows a normal distribution, and it is used when comparing means in situations where the standard deviation is known.

Key Formula

The Z-test statistic is calculated using the following formula:

$$Z = \frac{\underline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Where:

- $\underline{X}$ = Sample mean
- μ = Population mean (or the hypothesized mean)
- σ = Population standard deviation
- n = Sample size

Example: Z-Test for a Single Sample

Scenario: Suppose a teacher claims that the average score of students on a standardized test is 70. A researcher wants to test this claim by taking a sample of 50 students. The sample mean is found to be 72, with a known population standard deviation of 10.

Step 1: State the Hypotheses

- Null Hypotheses (H_0): The mean score of the students is 70. ($\mu = 70$)
- Alternative Hypotheses (H_1): The mean score of the students is not 70. ($\mu \neq 70$).

Step 2: Calculate the Z-Test Statistic Using the formula:

$$Z = \frac{72-70}{\frac{10}{\sqrt{50}}} = \frac{2}{\frac{10}{7.07}} = \frac{2}{1.414} \approx 1.41$$

Step 3: Determine the Critical Value and Make a Decision Assuming a significance level (α) of 0.05 for a two-tailed test, the critical value of Z is approximately ±1.96.

- Since the calculated Z value (1.41) is less than the critical value (1.96), we fail to reject the null hypothesis.

Conclusion: There is not enough evidence to reject the teacher's claim that the average score is 70.

This example illustrates how the Z-test is used to compare a sample mean to a population mean, allowing researchers to make inferences about the population based on sample data.

11.5.2 T Test

The T-test is a parametric statistical test used to determine whether there is a significant difference between the means of two groups. It is particularly useful when the sample size is small and the population standard deviation is unknown. There are different types of T-tests, depending on the research question and study design.

Types of T-Tests:

1. One-Sample T-Test: Compares the mean of a single sample to a known or hypothesized population mean.
2. Independent Samples T-Test (Two-Sample T-Test): Compares the means of two independent groups to determine if they are significantly different from each other.
3. Paired Samples T-Test (Dependent T-Test): Compares the means of two related groups, such as measurements taken from the same subjects at two different times.

Key Formula

For an Independent Samples T-Test, the T-test statistic is calculated using the following formula:

$$t = \frac{\underline{X}_1 - \underline{X}_2}{\sqrt{\frac{\delta_1^2}{\sqrt{n_1}} + \frac{\delta_2^2}{\sqrt{n_2}}}}$$

Where:

- $\underline{X}_1$ and $\underline{X}_2$ = Means of the two samples
- δ_1^2 and δ_2^2 = Variances of the two samples
- n_1 and n_2 = Sample sizes of the two groups

Example: Independent Samples T-Test

Scenario: Suppose a researcher wants to compare the effectiveness of two teaching methods. The researcher collects test scores from two groups of students: Group A (taught with method 1) and Group B (taught with Method 2). The test scores for both groups are as follows:

- Group A: n_1 = 10, $\underline{X}_1$ = 85, δ_1 = 5
- Group B: n_2 = 10, $\underline{X}_2$ = 80, δ_1= 6

Step 1: State the Hypotheses

- Null Hypotheses (H_0): There is no difference in the mean test scores between the two groups. $(\mu_1 \neq \mu_2)$
- Alternative Hypotheses (H_1): There is a difference in the mean test scores between the two groups. $(\mu_1 \neq \mu_2)$

Step 2: Calculate the T-Test Statistic

Using the formula:

$$t = \frac{85-80}{\sqrt{\frac{5^2}{10}+\frac{6^2}{10}}} = \frac{5}{\sqrt{\frac{25}{10}+\frac{36}{10}}} = \frac{5}{\sqrt{2.5+3.6}} = \frac{5}{\sqrt{6.1}} = \frac{5}{2.47} \approx 2.02$$

Step 3: Determine the Degrees of Freedom (df) For an independent samples T-test, the degrees of freedom are calculated as:

$$df = n_1 + n_2 - 2 = 10 + 10 - 2 = 18$$

Step 4: Compare the T-Statistic to the Critical Value

Assuming a significance level (α) of 0.05 for a two-tailed test, the critical value for df = 18 is approximately $\pm$2.101.

- Since the calculated t-value (2.02) is less than the critical value (2.101), we fail to reject the null hypothesis.

Conclusion: There is not enough evidence to conclude that there is a significant difference in the mean test scores between the two teaching methods.

This example demonstrates how the T-test is used to compare the means of two groups, allowing researchers to make inferences about the effectiveness of different treatments or interventions.

11.5.3 One-way ANOVA

One-way ANOVA (Analysis of Variance) is a statistical test used to determine whether there are any statistically significant differences between the means of three or more independent groups. It extends the two-sample t-test to more than two groups. The test assesses whether the variation between the group means is greater than the variation within each group, indicating whether at least one group mean is different from the others.

Key Formula

The F-statistic for One-Way ANOVA is calculated using the following formula:

$$F = \frac{Between-Group\ Variance}{Within-Group\ Variance}$$

Where:

- **Between-Group Variance**: Measures how much the group means deviate from the overall mean.
- **Within-Group Variance**: Measures the variability of observations within each group.

The Between-Group Variance (Mean Square Between, MSB) and Within-Group Variance (Mean Square Within, MSW) are calculated as follows:

$$MSB = \frac{SSB}{d\,f_{between}}$$

$$MSB = \frac{SSB}{d\,f_{within}}$$

Where:

- SSB = Sum of Squares Between
- SSB = Sum of Squares Within

- $d\,f_{between}$ = Degrees of Freedom Between Groups
- $d\,f_{within}$ = Degree of Freedom Within Groups

Example: One-Way ANOVA

Scenarios: Suppose a researcher wants to compare the effectiveness of three different diets on weight loss. They conduct an experiment with three groups, each following a different diet, and measure the weight loss after one month. The data collected is as follows:

- **Diet A**: n_1 = 5, weight loss (kg) = [3, 4, 5, 4, 6]
- **Diet B**: n_2 = 5, weight loss (kg) = [2, 2, 3, 2, 4]
- **Diet C**: n_3 = 5, weight loss (kg) = [5, 6, 6, 7, 8]

Step 1: State the Hypotheses

- Null Hypotheses (H_0): All group means are equal ($\mu_A = \mu_B = \mu_C$).
- Alternative Hypotheses (H_1): At least one group mean is different.

Step 2: Calculate the Group Means and Overall Mean

- **Mean for Diet A**: $\underline{X}_A = \frac{3+4+5+4+6}{5} = 4.4$
- **Mean for Diet B**: $\underline{X}_B = \frac{2+2+3+2+4}{5} = 2.6$
- **Mean for Diet C**: $\underline{X}_C = \frac{5+6+6+7+8}{5} = 6.4$
- **Overall Mean**: $\underline{X}_{overall} = \frac{(4.4 \times 5)+(2.6 \times 5)+(6.4 \times 5)}{5} = \frac{22+13+32}{15} = 2.733$

Step 3: Calculate the Sum of Squares

- **Sum of Squares Between (SSB):**

$$SSB = n_1(\underline{X}_A - \underline{X}_{overall})^2 + n_2(\underline{X}_B - \underline{X}_{overall})^2 + n_3(\underline{X}_C - \underline{X}_{overall})^2$$

$$SSB = 5 \times (4.4 - 4.27)^2 + 5 \times (2.6 - 4.27)^2 + 5 \times (6.4 - 4.27)^2$$

$$SSB = 5 \times (0.13)^2 + 5 \times (1.67)^2 + 5 \times (2.13)^2$$

$$= 5 \times 0.017 + 5 \times 2.80 + 5 \times 4.53 = 0.$$

- **Sum of Squares Within (SSW):**

$$SSW = \sum (\underline{X}_{ij} - \underline{X}_i)^2$$

$$SSW = \sum (weight\ loss\ for\ Diet\ A) - 4.4 + (weight\ loss\ for\ Diet\ B) - 2.6 + (weight\ loss\ for\ Diet\ C) - 6.4$$

$$SSW = ((3 - 4.4)^2 + (4 - 4.4)^2 + (5 - 4.4)^2 + (4 - 4.4)^2 + (6 - 4.4)^2 + ((2 - 2.6)^2 +$$

$$SSW = 2.0 + 0.2 + 0.4 + 0.2 + 2.8 + 0.4 + 0.4 + 0.2 + 0.4 + 2.8 + 0.4 + 0.2 + 0.2 + 0$$

Step 4: Calculate the Mean Squares

- **Mean Square Between (MSB)**:

$$MSB = \frac{SSB}{df_{between}} = \frac{36.73}{3-1} = 18.37$$

- **Mean Square Within (MSW)**:

$$MSB = \frac{SSB}{df_{within}} = \frac{16.7}{15-3} = 1.67$$

Step 5: Calculate the F-Statistic

$$F = \frac{MSB}{MSW} = \frac{18.37}{1.67} = 11.02$$

Step 6: compare the F-Statistic to the Critical Value

Assuming a significance level (α) of 0.05, the critical value for $df_{between} = 2 \; and \; df_{within} = 12$ can be found in F-distribution tables. It is approximately 3.89.

- Since the calculated F-value (11.02) is greater than the critical value (3.89), we reject the null hypothesis.

Conclusion: There is a significant difference in the mean weight loss between at least one of the diet groups.

This example demonstrates how One-Way ANOVA is used to compare the means of multiple groups and determine if there are significant differences between them.

Check Your Progress A:

1. What is a t-test? When is it used, and for what purpose(s)? Explain by means of examples.

....

....

....

....

2. What is the significance level?

....

....

....

....

3. Distinguish between the following:

i) Null hypothesis and alternative hypothesis;

ii) One-tailed test and two-tailed test;

iii) Type I error and Type II error;

iv) Acceptance region and rejection region;

4. Choose the correct option

   i) What is the primary purpose of hypotheses formulation?

      a) To collect raw data

      b) To make assumption without testing

      c) To make a clear, testable statement about relationship between variable.

      d) To summarize statistical findings

   ii) Which statistical test is commonly used to compare mean of two independent groups?

      a) Chi-square test

      b) ANNOVA

      c) T-Test

      d) Regression analysis

   iii) What does a p-value indicate in hypotheses testing?

      a) Probability that null hypothesis is false

      b) Probability of observing data if null hypothesis true

      c) Confidence interval of test

      d) mean difference between two data set.

11.6 NON-PARAMETRIC TEST

Non-parametric tests are statistical methods used for analysing data that do not meet the assumptions required for parametric tests, such as normal distribution or homogeneity of variances. These tests are particularly valuable when dealing with ordinal data, non-normally distributed interval data, or when sample sizes are small. Unlike parametric tests, non-parametric tests do not rely on assumptions about the distribution of the data, making them robust and versatile tools in statistical analysis.

One of the primary characteristics of non-parametric tests is their distribution-free nature. They do not assume a specific distribution for the data, which makes them suitable for datasets that deviate from normality. This quality also makes them less sensitive to outliers and skewed distributions compared to parametric tests.

Several commonly used non-parametric tests serve various purposes:

1. **Mann-Whitney U Test:** This test compares differences between two independent groups. For example, it can be used to compare the test scores of students taught using two different methods when the scores do not follow a normal distribution.

2. **Wilcoxon Signed-Rank Test:** This test is used for comparing two related samples or repeated measurements on a single sample. It is often employed to assess the impact of a treatment by comparing measurements before and after the intervention within the same group.

3. **Kruskal-Wallis H Test:** An extension of the Mann-Whitney U test, the Kruskal-Wallis H test compares more than two independent groups. It is useful for analysing differences in satisfaction levels among several groups when the data does not follow a normal distribution.

4. **Friedman Test:** This test is used for comparing more than two related samples or repeated measurements. It is commonly applied in experiments where the same subjects are measured under different conditions, such as evaluating the effectiveness of multiple treatments over time.

5. **Chi-Square Test:** The Chi-Square test is used for categorical data to assess independence or goodness-of-fit. For example, it can test whether there is an association between gender and voting preference. We will study about the Z-test in detail in coming head 11.6.1.

6. **Spearman's Rank Correlation Coefficient:** This test measures the strength and direction of the association between two ranked variables. It is useful for examining relationships between variables where the data is ordinal or not normally distributed.

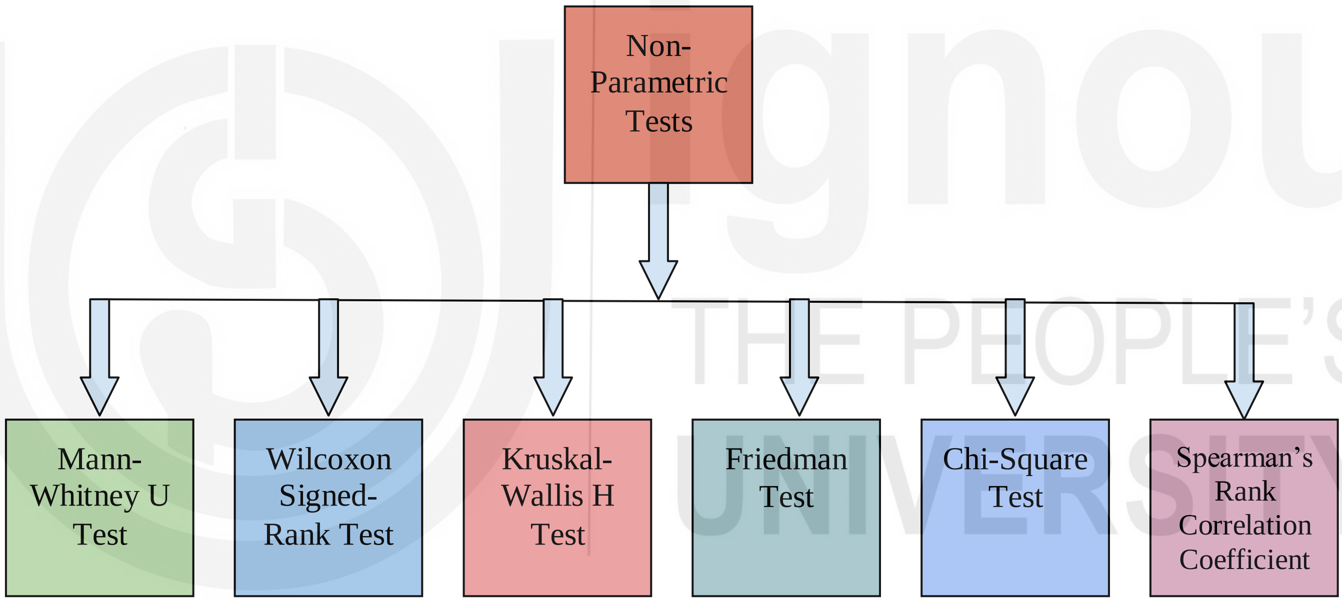


Figure 11.4: Types of Non-Parametric Tests

Non-parametric tests are especially valuable in situations where traditional parametric methods are not applicable due to violations of assumptions or when dealing with non-quantitative data. They offer flexible and robust alternatives for drawing meaningful conclusions from data, ensuring that researchers can analyse and interpret results even when parametric assumptions are not met.

11.6.1 Chi square Test

Chi-square is an important non-parametric test and as such no rigid assumptions are necessary in respect of the type of population. We require only the degrees of freedom (implicitly of course the size of the sample) for using this test. As a non-parametric test, chi-square can be used as follows:

i) as a test of goodness of fit and

ii) as a test of independence.

Chi-Square Test for Independence

As a test of independence, χ^2 test enables us to explain whether or not two attributes are associated. For instance, we may be interested in knowing whether a new medicine is effective in controlling fever or not, χ^2 test will helps us in deciding this issue. In such a situation, we proceed with the null hypothesis that the two attributes (viz., new medicine and control of fever) are independent which means that new medicine is not effective in controlling fever. On this basis we first calculate the expected frequencies and then work out the value of χ^2 . If the calculated value of χ^2 is less than the table value at a certain level of significance for given degrees of freedom, we conclude that null hypothesis stands which means that the two attributes are independent or not associated (i.e., the new medicine is not effective in controlling the fever). But if the calculated value of χ^2 is greater than its table value, our inference then would be that null hypothesis does not hold good which means the two attributes are associated and the association is not because of some chance factor but it exists in reality (i.e., the new medicine is effective in controlling the fever and as such may be prescribed). It may, however, be stated here that χ^2 is not a measure of the degree of relationship or the form of relationship between two attributes, but is simply a technique of judging the significance of such association or relationship between two attributes. In order that we may apply the chi-square test either as a test of goodness of fit or as a test to judge the significance of association between attributes, it is necessary that the observed as well as theoretical or expected frequencies must be grouped in the same way and the theoretical distribution must be adjusted to give the same total frequency as we find in case of observed distribution. χ^2 is then calculated as follows:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where.,

Oij = observed frequency of the cell in ith row and jth column.

Eij = expected frequency of the cell in ith row and jth column.

If two distributions (observed and theoretical) are exactly alike, $\chi 2 = 0$; but generally due to sampling errors, $\chi 2$ is not equal to zero and as such we must know the sampling distribution of $\chi 2$ so that we may find the probability of an observed $\chi 2$ being given by a random sample from the hypothetical universe. Instead of working out the probabilities, we can use ready table which gives probabilities for given values of $\chi 2$. Whether or not a calculated value of $\chi 2$ is significant can be ascertained by looking at the tabulated values of $\chi 2$ for given degrees of freedom at a certain level of significance. If the calculated value of $\chi 2$ is equal to or exceeds the table value, the difference between the observed and expected frequencies is taken as significant, but if the table value is more than the calculated value of $\chi 2$, then the difference is considered as insignificant i.e., considered to have arisen as a result of chance and as such can be ignored. As already stated, degrees of freedom play an important part in using the chi-square distribution and the test based on it, one must correctly determine the degrees of freedom. If there are 10

frequency classes and there is one independent constraint, then there are (10 – 1) = 9 degrees of freedom. Thus, if 'n' is the number of groups and one constraint is placed by making the totals of observed and expected frequencies equal, the degrees of freedom (d.f.) would be equal to (n – 1). In the case of a contingency table (i.e., a table with 2 columns and 2 rows or a table with two columns and more than two rows or a table with two rows but more than two columns or a table with more than two rows and more than two columns), the d.f. is worked out as follows: d.f. = (c – 1) (r – 1) where 'c' means the number of columns and 'r' means the number of rows.

Mathematical Equation for Chi-Square Test for Independence

The formula for the Chi-Square statistic (χ^2) is:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where:

- Oi = Observed frequency for category i
- Ei = Expected frequency for category i
- $\sum$ = Summation over all categories

The expected frequency for each cell in a contingency table is calculated as:

$$E_i = \frac{(\text{Row Total} \times \text{Column Total})}{\text{Grand Total}}$$

Example: Chi-Square Test for Independence

Scenario: Suppose we want to test whether there is an association between gender (Male, Female) and preference for a product (Product A, Product B).

We collect data from a sample of 100 people:

Gender	**Product A**	**Product B**	**Row Total**
Male	30	20	50
Female	10	40	50
Column Total	40	60	100

Step 1: State the Hypotheses

- Null Hypotheses (H0): There is no association between gender and product preference.
- Alternative Hypotheses (H1): There is an association between gender and product preference.

Step 2: Calculate the Expected Frequencies

Using the formula for expected frequency:

$$E_{11} = \frac{(Row\ Total\ for\ Male\ \times Column\ Total\ for\ Product\ A)}{Grand\ Total}$$

$$= \frac{(50 \times 40)}{100} = 20$$

Similarly, calculate E_{12}, E_{21}, and E_{22}:

- $E_{12} = \frac{(50\times60)}{100} = 30$
- $E_{21} = \frac{(50\times40)}{100} = 20$
- $E_{22} = \frac{(50\times60)}{100} = 30$

Expected Frequency Table;

Gender	Product A	Product B
Male	20	30
Female	20	30

Step 3: Calculate the Chi-Square Statistic

Using the formula:

$$x^2 = \frac{(30-20)^2}{20} + \frac{(20-30)^2}{30} + \frac{(10-20)^2}{20} + \frac{(40-30)^2}{30}$$

$$x^2 = \frac{100}{20} + \frac{100}{30} + \frac{100}{20} + \frac{100}{30} = 5 + 3.33 + 5 + 3.33 = 16.66$$

Step 4: Determine the Degree of Freedom (df) and the Critical Value

Degrees of freedom df for a Chi-Square test is calculated as:

$$df = (r-1) \times (c-1)$$

Where r is the number of rows and c is the number of columns. In this case:

$$df = (2-1) \times (2-1) = 1$$

Step 5: Compare the Chi-Square statistic to the Critical Value

Using a Chi-Square distribution table, find the critical value for $df = 1$ and a significance level of $\alpha = 0.05$), which is approximately 3.84.

Since the calculated x^2 value (16.66) is greater than the critical value (3.84), we reject the null hypothesis.

Conclusion: There is significant evidence to suggest that there is an association between gender and product preference.

Chi-Square Goodness-of-Fit Test

As a test of goodness of fit, $\chi 2$ test enables us to see how well does the assumed theoretical distribution (such as Binomial distribution, Poisson distribution or Normal distribution) fit to the observed data. When some theoretical distribution is fitted to the given data, we are always interested in knowing as to how well this distribution fits with the observed data. The chi-square test can give answer to this. If the calculated value of $\chi 2$ is less than the table value at a certain level of significance, the fit is considered to be a good one which means that the divergence between the observed and

expected frequencies is attributable to fluctuations of sampling. But if the calculated value of χ2 is greater than its table value, the fit is not considered to be a good one.

Chi-Square Goodness-of-Fit Test

Mathematical Equation

The formula remains the same as for the Chi-Square test for independence:

$$x^2 \sum \frac{(O_i - E_i)^2}{E_i}$$

Where O_i are the observed frequencies, and E_i are the expected frequencies based on a specific distribution.

Example: Chi-Square Goodness-of-Fit Test

Scenario: Suppose we want to test whether a die is fair. We roll the die 60 times and record the frequency of each outcome.

Outcome	1	2	3	4	5	6
Observed Frequency	8	12	10	9	11	10

Step 1: State the Hypotheses

- Null Hypotheses (H_0): The die is fair, so each outcome has an equal probability of $\frac{1}{6}$.
- Alternative Hypotheses (H_1): The die is not fair, so the outcomes do not have equal probabilities.

Step 2: Calculate the Expected Frequencies

If the die is fair, the expected frequency for each outcome is:

$$E_i = \frac{Total\ Rolls}{Number\ of\ Outcomes} = \frac{60}{6} = 10$$

Step 3: Calculate the Chi-Square statistic

$$x^2 = \frac{(8-10)^2}{10} + \frac{(12-10)^2}{10} + \frac{(10-10)^2}{10} + \frac{(9-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(10-10)^2}{10}$$

$$x^2 = \frac{4}{10} + \frac{4}{10} + \frac{0}{10} + \frac{1}{10} + \frac{1}{10} + \frac{0}{10} = 0.4 + 0.4 + 0 + 0.1 + 0.1 + 0 = 1.0$$

Step 4: Determine the Degrees of Freedom (df) and the Critical Value

Degree of freedom $d\,f$ is calculated as:

$$d\,f = k - 1$$

Where k is the number of categories (outcomes). In this case:

$$df = 6 - 1 = 5$$

Step 5: Compare the Chi-Square Statistic to the Critical Value

Using a Chi-Square distribution table, find the critical value for $df = 5$ and a significance level of $\alpha = 0.05$, which is approximately 11.07.

Since the calculated x^2 value (1.0) is less than the critical value (11.07), we fail to reject the null hypothesis.

Conclusion: There is no significant evidence to suggest that the die is unfair. It appears to be fair based on the data collected.

These examples demonstrate how the Chi-Square test is used to assess relationships between categorical variables and the goodness of fit for observed data.

CONDITIONS FOR THE APPLICATION OF χ2 TEST

The following conditions should be satisfied before χ2 test can be applied:

i) Observations recorded and used are collected on a random basis.

ii) All the items in the sample must be independent.

iii) No group should contain very few items, say less than 10. In case where the frequencies are less than 10, regrouping is done by combining the frequencies of adjoining groups so that the new frequencies become greater than 10. Some statisticians take this number as 5, but 10 is regarded as better by most of the statisticians.

iv) The overall number of items must also be reasonably large. It should normally be at least 50, howsoever small the number of groups may be.

v) The constraints must be linear. Constraints which involve linear equations in the cell frequencies of a contingency table (i.e., equations containing no squares or higher powers of the frequencies) are known are known as linear constraints.

Here are the main differences between parametric and non-parametric tests in more detail which you can able to understand clearly from the below table;

Table 11.1: Difference between Parametric and Non-Parametric Tests

Criterion	**Parametric**	**Non-Parametric**
Population	A proper understanding of the population is available	Not much information about the population is available
Assumptions	Several assumptions are regarding the population. Incorrect results are provided if assumptions are not filled	No assumptions are made regarding the population.
Distribution	The distribution of the population is often required to be normal	Do not require the population to be normal; it can be arbitrary
Sample size	Require sample size to be over 30	Can work with small samples

Interpretability	Are easy to interpret	Are difficult to interpret
Implementation	Are difficult to implement	Are easy to implement
Reliability	The output is more powerful/reliable	Are less powerful/reliable
Type of variable	Works with continuous/ quantitative variable	Works with continuous/ quantitative variable as well as categorical or discrete variable
Central tendency	Measurement of central tendency is typical done using mean	Measurement of central tendency is generally done using median
Outliers	Affected by Outliers	Less affected by Outliers
Null hypothesis	More acute	Incorrectly rejects null hypothesis; less accurate
Examples	z-test, t-test, Anova, f-test, Pearson coefficient of correlation	One sample KS test, Wilcoxon singed rank test, Mann whitney u-test, Wilcoxon rank sum test, Krusal wallis test, Spearman rank coefficient, Chi square test for independence

11.7 CORRELATION

Correlation is a statistical technique used to measure and describe the strength and direction of the relationship between two or more variables. It quantifies how changes in one variable are associated with changes in another. Understanding correlation helps in predicting and identifying relationships between variables in various fields such as science, economics, and social sciences.

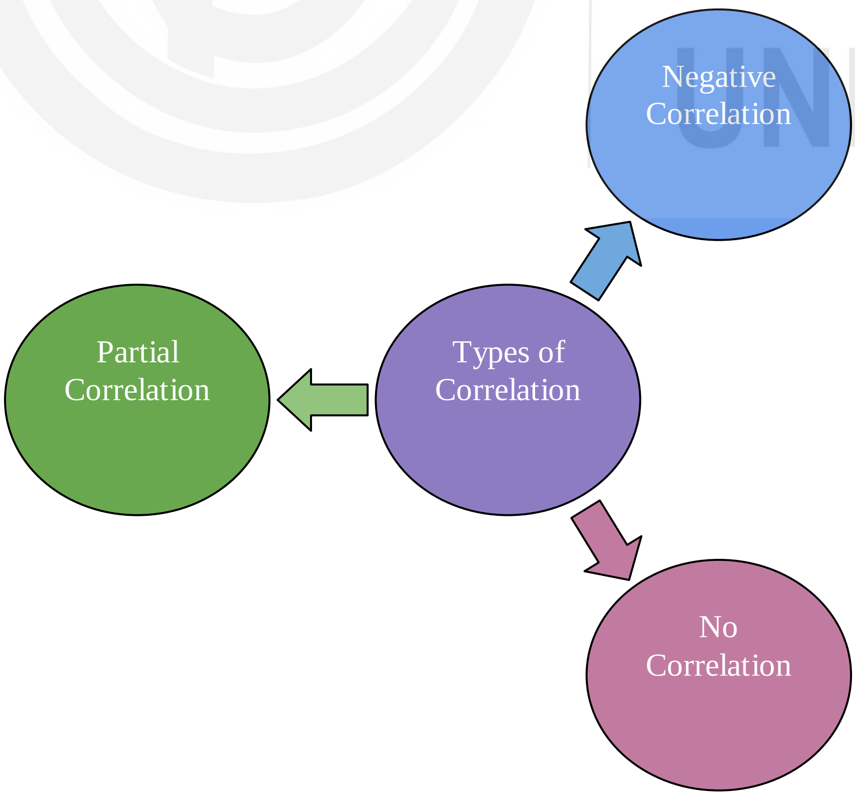


Figure 11.5: Types of Correlation

Types of Correlation:

1. **Positive Correlation:**

   - **Definition:** Occurs when an increase in one variable is associated with an increase in another variable. Conversely, a decrease in one variable corresponds to a decrease in the other.
   - **Example:** The relationship between hours studied and exam scores. Generally, as the number of hours studied increases, exam scores also increase.

2. **Negative Correlation:**

   - **Definition:** Occurs when an increase in one variable is associated with a decrease in another variable, and vice versa.
   - **Example:** The relationship between the number of hours spent on leisure activities and the number of hours spent working. As leisure time increases, work time typically decreases.

3. **No Correlation:**

   - **Definition:** Occurs when there is no apparent relationship between two variables. Changes in one variable do not predict changes in the other.
   - **Example:** The relationship between shoe size and intelligence. Generally, there is no meaningful correlation between these two variables.

Measures of Correlation:

1. **Pearson Correlation Coefficient (r):**

   - **Definition:** Measures the linear relationship between two continuous variables. It ranges from -1 to 1.
     - 1: Perfect positive linear relationship.
     - -1: Perfect negative linear relationship.
     - 0: No linear relationship.

   Formula:

$$r = \frac{\Sigma \quad (X - \underline{X})(Y - \underline{Y})}{\sqrt{\Sigma \quad (X - \underline{X})^2 \Sigma \quad (Y - \underline{Y})^2}}$$

   - **Example:** If the Pearson correlation coefficient between height and weight is 0.85, it indicates a strong positive linear relationship.

2. **Spearman's Rank Correlation Coefficient (ρ):**

   - **Definition**: Measures the strength and direction of association between two ranked variables. It is used for ordinal data or when the assumptions of Pearson correlation are not met.
   - **Formula**:

$$\rho = 1 - \frac{6 \Sigma \quad d_i^2}{n(n^2 - 1)}$$

Where d_i is the difference between the ranks of each pair, and n is the number of pairs.

- **Example**: If the Spearman's rank correlation coefficient between two sets of ranked data (e.g., preferences for two different products) is 0.7, it indicates a strong positive rank-order correlation.

3. **Kendall's Tau (τ):**

- **Definition**: Measures the strength and direction of association between two variables based on the ranks. It is more robust than Pearson and Spearman's correlation for small sample sizes and ties.
- **Formula:**

$$\tau = \frac{(C - D)}{\sqrt{(C + D + T_1)(C + D + T_2)}}$$

Where C is the number of concordant pairs, D is the number of discordant pairs, and T_1 and T_2 are the number of tied pairs in each variable.

- **Example:** If Kendall's Tau between two variables is 0.4, it suggests a moderately positive correlation between the variables.

Applications:

Correlation analysis is widely used to:

- **Predict outcomes:** Understanding how changes in one variable affect another.
- **Identify Relationships:** Discovering potential associations between variables.
- **Inform Decisions:** Assist in making data-driven decisions based on the relationships between variables.

By analyzing correlation, researchers can uncover patterns and relationships that inform their understanding and guide further investigation.

Let's work through a numerical example to calculate the Karl Pearson coefficient of correlation (Pearson's r).

Example Scenario:

Suppose we want to find the correlation between the number of hours studied and the scores obtained on a test by 5 students.

Data:

Student	Hours Studied(X)	Test Score (Y)
1	2	50
2	4	55
3	6	60
4	8	65
5	10	70

Steps to Calculate Pearson's Correlation Coefficient

1. **Calculate the Means:**

- Mean of $X(\underline{X})$:

$$\underline{X} = \frac{2+4+6+8+10}{5} = \frac{30}{5} = 6$$

- Mean of $Y(\underline{Y})$:

$$\underline{Y} = \frac{50+55+60+65+70}{5} = \frac{300}{5} = 60$$

2. **Calculate the Deviations from the Mean and Their Products:**

Student	X	Y	$X-\underline{X}$	$Y-\underline{Y}$	$(X-\underline{X}) \times (Y-\underline{Y})$
1	2	50	-4	-10	40
2	4	55	-2	-5	10
3	6	60	0	0	0
4	8	65	2	5	10
5	10	70	4	10	40

- Sum of Products:

$$\sum (X-\underline{X}) \times (Y-\underline{Y}) = 40+10+0+10+40 = 100$$

3. **Calculate the Sum of Squares:**

- Sum of Squares for $X(\sum (X-\underline{X})^2)$:

$$\sum \left((X-\underline{X})^2\right) = (-4)^2 + (-2)^2 + 0^2 + 2^2 + 4^2 = 16+4+0+4+6$$
$$= 40$$

Sum of squares for $Y(\sum (Y-\underline{Y})^2)$:

$$\sum \left((Y-\underline{Y})^2\right) = (-10)^2 + (-5)^2 + 0^2 + 5^2 + 10^2$$

$$= 100+25+0+25+100 = 250$$

4. **Calculate the Pearson Correlation Coefficient *r*:**

$$r = \frac{\sum (X-\underline{X}) \times (Y-\underline{Y})}{\sqrt{\sum (X-\underline{X})^2 \times \sum (Y-\underline{Y})^2}}$$

$$r = \frac{100}{\sqrt{40 \times 250}} = \frac{100}{\sqrt{10000}} = \frac{100}{100} = 1$$

Interpretation:

- ***r* = 1:** This indicates a perfect positive linear relationship between the number of hours studied and the test scores. As the number of hours studied increases, the test scores increase in a perfectly linear manner.

Let's work through a numerical example for Spearman's Rank correlation Coefficient (ρ).

Example Scenario:

Suppose we want to find the correlation between the rankings of five students in two different subjects, say math and science.

Data:

Student	Math Rank (X)	Science Rank (Y)
1	1	2
2	2	1
3	3	4
4	4	3
5	5	5

Steps to Calculate Spearman's Rank Correlation Coefficient:

1. **Calculate the Difference Between Ranks (d) and Squared Differences (d^2):**

Student	*X*	*Y*	*d = X- Y*	d^2
1	*1*	*2*	*-1*	*1*
2	*2*	*1*	*1*	*1*
3	*3*	*4*	*-1*	*1*
4	*4*	*3*	*1*	*1*
5	*5*	*5*	*0*	*0*

- Sum of d^2 :

$$\sum d^2 = 1 + 1 + 1 + 1 + 0 = 4$$

2. Apply the Spearman's Rank Correlation Formula:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

Where n is the number of observations (students).

- In this case, $n = 5$.

$$\rho = 1 - \frac{6 \times 4}{5(5^2 - 1)}$$

- Calculate $5^2 - 1$;

$$5^2 - 1 = 25 - 1 = 24$$

- Substitute values;

$$\rho = 1 - \frac{24}{120} = 1 - 0.2 = 0.8$$

Interpretation:

- Spearman's Rank Correlation Coefficient $\rho = 0.8$: This indicates a strong positive relationship between the rankings in Math and Science. As students rank higher in Math, they also tend to rank higher in science, and vice versa.

11.8 REGRESSION

Regression analysis is a statistical method used to understand and quantify the relationship between a dependent variable and one or more independent variables. It helps in predicting the value of the dependent variable based on the values of the independent variables. Regression models are essential in various fields such as economics, biology, engineering, management and social sciences for making predictions and understanding relationships.

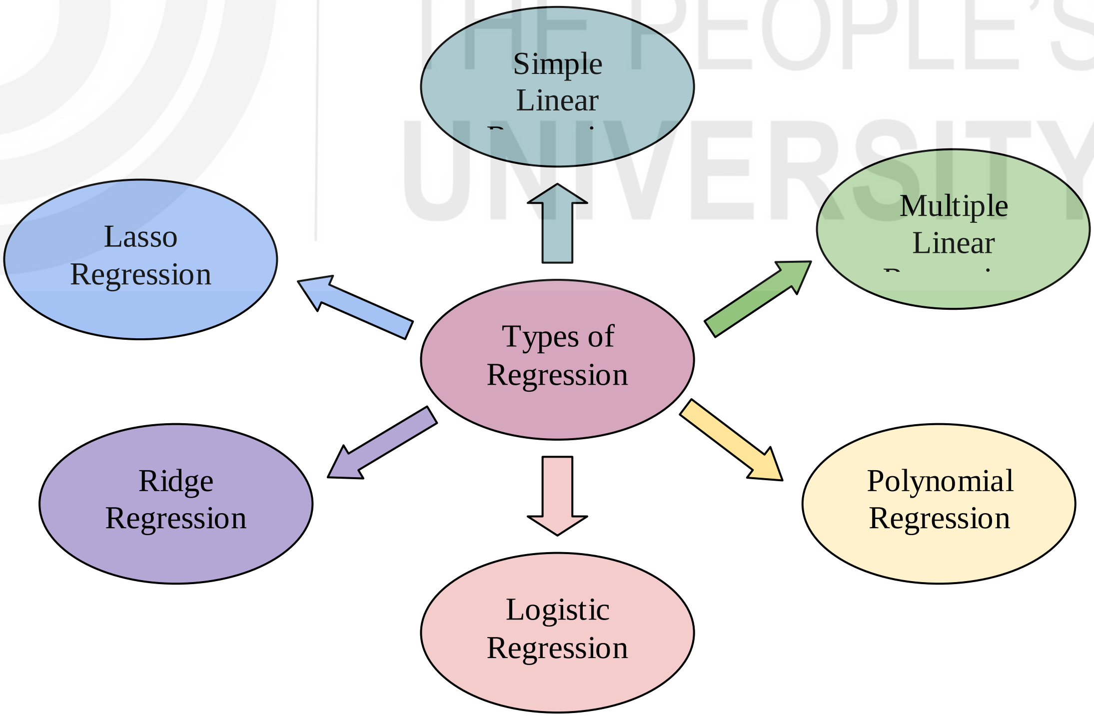


Figure 11.6: Types of Regression

Types of Regression:

1. **Simple Linear Regression:**

   - **Definition:** Examines the relationship between a single independent variable and a dependent variable by fitting a straight line to the data.
   - **Equation:** $Y = \beta_0 + \beta_1 X + \epsilon$

   Where Y is the dependent variable, X is the independent variable, β_0 is the y-intercept, β_1 is the slope of the line, and ϵ is the error term.

   - **Example:** Predicting a person's weight based on their height.

2. **Multiple Linear Regression:**

   - **Definition:** Extends simple linear regression to include multiple independent variables. It models the relationship between a dependent variable and two or more independent variables.
   - **Equation:** $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon$

     Where X_1, X_2, ...X_k are the independent variables, and $\beta_1, \beta_2, \ldots .\beta_k$ are the coefficients for these variables.
   - **Example:** Predicting a person's salary based on their years of experience, education level, and job position.

3. **Polynomial Regression:**

   - **Definition:** A form of regression analysis in which the relationship between the independent variable and the dependent variable is modeled as an nth-degree polynomial.
   - **Equation:** $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \cdots + \beta_n X^n + \epsilon$
   - **Example:** Modeling the trajectory of a projectile, where the relationship between time and distance traveled is not linear.

4. **Logistic Regression:**

   - **Definition:** Used when the dependent variable is categorical (often binary). It estimates the probability of a binary outcome based on one or more independent variables.
   - **Equation:**

     $$logit(P) = In\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

     Where P is the probability of the event occurring.
   - **Example:** Predicting whether an email is spam (yes/no) based on certain features of the email.

5. **Ridge Regression:**

   - **Definition:** A type of linear regression that includes a regularization term to prevent overfitting by penalizing large coefficients.

- **Equation:**

$$Minimize\ (\sum\ (Y_i - \hat{Y}_i) + \lambda \sum\ \beta_j^2)$$

Where λ is a regularization parameter.

- **Example:** Predicting house prices with many features, where some features may be highly correlated.

6. **Lasso Regression:**

- **Definition:** Similar to ridge regression, but it includes an L1 penalty that can shrink some coefficients to zero, effectively performing feature selection.
- **Equation:**

$$Minimize\ (\sum\ (Y_i - \hat{Y}_i)^2 + \lambda \sum\ |\beta_j|)$$

- **Example:** Building a predictive model for stock prices where feature selection is crucial due to the high dimensionality of the data.

Applications:

- **Prediction:** Making forecasts based on historical data.
- **Understanding Relationships:** Identifying and quantifying relationships between variables.
- **Decision Making:** Informing decisions based on model outputs and predictions.

Regression analysis provides a powerful framework for analyzing relationships between variables and making predictions based on those relationships.

Check Your Progress B:

1. What does a correlation coefficient of 0.85 indicate about the relationship between two variables?

   ....

   ....

   ....

   ....

   ....

2. Write short notes on the following:

   i) Non-Parametric Tests;

   ii) Chi-square as a test of 'goodness of fit';

   iii) Correlation;

   iv) Regression Analysis.

11.9 LET US SUM UP

Hypotheses is usually considered as the principal instrument in research. Its main function is to suggest new experiments and observations. Hypotheses testing enables us to make probability statements about population parameter(s). The hypothesis may not be proved absolutely, but in practice it is accepted if it has withstood a critical testing. Formulating a hypothesis is a systematic process that involves several key steps to ensure that the hypothesis is clear, testable, and relevant to the research question. While hypothesis testing is a critical process in statistics that allows researchers to make inferences about a population based on sample data. Parametric tests such as T-test, Z-test, ANOVA, F-test are statistical tests that assume the underlying data follows a certain distribution, typically a normal distribution, and rely on parameters such as the mean and standard deviation to make inferences about the population. Non-parametric tests such as Mann-Whitney U Test, Wilcoxon Signed-Rank Test, Kruskal-Wallis H Test, Chi-Square Test are statistical methods used for analysing data that do not meet the assumptions required for parametric tests, such as normal distribution or homogeneity of variances. These tests are particularly valuable when dealing with ordinal data, non-normally distributed interval data, or when sample sizes are small. Correlation is a statistical technique used to measure and describe the strength and direction of the relationship between two or more variables. Regression analysis is a statistical method used to understand and quantify the relationship between a dependent variable and one or more independent variables. It helps in predicting the value of the dependent variable based on hypothesis testing enables us to make probability statements about population parameter(s). Regression analysis provides a powerful framework for analyzing relationships between variables and making predictions based on those relationships. The hypothesis may not be proved absolutely, but in practice it is accepted if it has withstood a critical testing.

11.10 KEYWORDS

Correlation: Correlation is a statistical technique used to measure and describe the strength and direction of the relationship between two or more variables.

Chi-Square Test: The Chi-Square test is a statistical procedure for determining the difference between observed and expected data.

Hypotheses: A hypothesis is an assumption, an idea that is proposed for the sake of argument so that it can be tested to see if it might be true.

Non-Parametric Tests: Non-parametric tests are statistical methods used for analyzing data that do not meet the assumptions required for parametric tests, such as normal distribution or homogeneity of variances.

Parametric Tests: Parametric tests are statistical tests that assume the underlying data follows a certain distribution, typically a normal distribution, and rely on parameters such as the mean and standard deviation to make inferences about the population.

Regression: Regression analysis is a statistical method used to understand and quantify the relationship between a dependent variable and one or more independent variables.

T-Test: The T-test is a parametric statistical test used to determine whether there is a significant difference between the means of two groups.

Z-Test: The Z-test is a parametric statistical test used to determine whether there is a significant difference between sample and population means or between two sample means.

11.11 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress A

4. (i) c (ii) c (iii) b

11.12 TERMINAL QUESTIONS

1. A sample of 400 male students is found to have a mean height of 67.47 inches. Can it be reasonably regarded as a sample from a large population with a mean height of 67.39 inches and a standard deviation of 1.30 inches? Test at 5% level of significance.
2. The procedure of testing hypothesis requires a researcher to adopt several steps. Describe in brief all such steps.
3. Briefly describe the important parametric tests used in the context of testing hypotheses. How do such tests differ from non-parametric tests? Explain.
4. A sample of 10 is drawn randomly from a certain population. The sum of the squared deviations from the mean of the given sample is 50. Test the hypothesis that the variance of the population is 5 at the 5 percent level of significance.
5. What is Chi-square text? Explain its significance in statistical analysis.
6. Briefly describe the different non-parametric tests, explaining the significance of each such test.
7. A company finds a correlation of -0.75 between customer complaints and customer satisfaction scores. How should the company interpret this result?
8. What is the difference between simple linear regression and multiple regression? Also, explain the role of the slope and intercept in a regression equation.

11.13 FURTHER READINGS

The following textbooks may be used for more in-depth study on the topics dealt with in this unit:

1. “Research Methodology: Methods and Techniques” by C.R. Kothari
2. “Business Research Methods” by H.K Dangri and Shruti Dewan
3. Raftery, A. E., Gilks, W. R., Richardson, S., & Spiegelhalter, D. (1995). Hypotheses testing and model. *Markov chain Monte Carlo in practice*, *1*, 165-87.

BLOCK 4
TECHNOLOGY IN RESEARCH AND REPORT WRITING

BLOCK 4 TECHNOLOGY IN RESEARCH AND REPORT WRITING

This block introduces learners to the transformative role of technology in research, which has made research processes more efficient, precise, and innovative. Technology serves as a catalyst, enabling researchers to push the boundaries of knowledge and discovery. This block explores the use of research software such as SPSS, Microsoft Office, R software, and Tableau, alongside citation metrics and reference management tools. It also addresses ethical considerations in research, emphasizing the importance of integrity and accountability. Finally, the block delves into the final stage of the research process—writing a research report—by discussing its types, preparation stages, and the structure of an effective report.

The block, titled **"Technology in Research and Report Writing,"** consists of four units, as detailed below:

Unit 12: Technology Usage in Research

This unit familiarizes learners with the various technological advancements that support and enhance research processes. It highlights the role of business research platforms, robotics, and automation as integral components of modern research. Additionally, it introduces emerging technologies such as cloud computing, augmented reality (AR), virtual reality (VR), and artificial intelligence (AI), emphasizing their growing significance in conducting research.

Unit 13: Software Application in Research

This unit focuses on the usage and application of various research software that facilitate the research process. Learners will gain insights into popular tools like SPSS, Microsoft Office, R software, and Tableau. The unit also covers essential research tools, including citation metrics and reference management systems, helping learners streamline their work and maintain accuracy in their studies.

Unit 14: Ethical Issues in Research

This unit emphasizes the importance of ethics in research and explores the ethical challenges that researchers often face. Learners will understand what research ethics entails, why it is crucial, and how ethical principles can guide them in maintaining integrity throughout the research journey. The unit also discusses common ethical issues and ways to address them effectively.

Unit 15: Report Writing

This final unit focuses on the last stage of the research process—writing the research report. It explains the different types of reports, the stages involved in preparing them, and the characteristics of a well-structured research report. Learners will become familiar with the components of a report, including prefatory items, the body, and terminal items, enabling them to create clear, comprehensive, and impactful reports.

UNIT 12 TECHNOLOGY USAGE IN RESEARCH

Structure

12.0 OBJECTIVES

After studying this unit, you will be able to:

- Explore the role of technology in enhancing the research process;
- Familiarize with advanced tools like cloud computing, VR, AR, and collaboration platforms;

- Know different AI tools used in Research; and
- Improve research efficiency and collaboration through modern technological tools.

12.1 INTRODUCTION

In the preceding unit, we investigated the process of formulating and testing hypotheses, which offered a methodical approach to validating research questions. Building on that foundation, this unit focuses on the transformative role of technology in enhancing the research process.

Technology is all around us. It's not just the devices we use daily, but the tools, systems, and innovations that make tasks faster, smarter, and more efficient. In research, technology opens doors to new possibilities. It allows us to analyze vast amounts of data, collaborate with peers across the globe, and present findings in dynamic ways. From simple tools like spreadsheets and survey software to advanced systems like data visualization platforms and simulations, technology has made research more accessible and impactful. Contemporary trends in research technology highlight the growing use of automation, big data analytics, and cloud computing. These advancements enable researchers to process complex datasets, collaborate seamlessly across geographies, and derive insights with unprecedented accuracy. One of the most impactful innovations in this landscape is Artificial Intelligence (AI). AI technologies, such as machine learning algorithms, natural language processing, and predictive modeling, are revolutionizing research by automating repetitive tasks, identifying patterns, and providing real-time analytics.

For instance, AI-powered tools like Chat GPT, Elicit.AI, and Sci Space can assist in literature reviews by analyzing thousands of articles within minutes, flagging relevant studies, and even summarizing findings. AI is also advancing fields like sciences, social sciences, and behavioral research by enabling researchers to delve deeper into intricate patterns and relationships.

In this unit, we will explore how technology is integrated into various stages of research. By understanding these tools, researchers can enhance their efficiency, and paving the way for innovation in their respective fields.

12.2 ABOUT TECHNOLOGY

Technology is an integral part of our daily lives, shaping how we communicate, work, and solve problems. At its core, technology refers to tools, systems, and processes created to improve efficiency, simplify tasks, and solve complex challenges. It spans a wide range of applications; from the simple tools we use for everyday tasks to advanced systems powered by artificial intelligence and robotics.

In the context of research, technology has evolved from being a support tool to becoming a driving force behind innovation. It provides researchers with the ability to collect, analyze, and interpret data more accurately and

efficiently. For instance, the use of specialized software like "Smart PLS", "AMOS", and "R" for statistical analysis reduces the time required for calculations while improving precision. Similarly, technologies like cloud computing allow researchers to store and access large datasets, making collaboration across the globe seamless.

The adaptability of technology also extends to research methodologies. Remote sensing, simulations, and digital surveys are examples of how technology adapts to the needs of various fields. In the previous unit of this course Review of Literature you have learned how digital libraries and online repositories like Sodh Ganga have revolutionized access to academic resources, breaking barriers of geography and time.

A key development in recent years is the rise of artificial intelligence (AI) in research. AI is a branch of technology that enables machines to learn, think, and make decisions. Its applications in research are vast—analyzing trends, predicting outcomes, or even automating repetitive tasks like formatting references. People are using AI to write a research paper. Software like "Mandley", "Zottero", and "Evernote" are used by researchers' for in-text citations and references.

As technology continues to advance, it has become more than a tool; it is a partner in exploration and discovery. By leveraging technology effectively, researchers can not only save time and resources but also push the boundaries of what we know and understand.

12.3 USAGE OF TECHNOLOGY IN RESEARCH

Technology has become a foundation of modern research, enabling tasks to be completed in minutes that once demanded extensive manual effort and attention. Machines work faster than humans in processing information, can handle complex computations, analyze vast datasets, and perform repetitive tasks efficiently, and ensure precision and consistency. By automating these processes, technology not only saves time but also significantly reduces costs, making high-quality research accessible to a broader range of institutions and researchers.

For instance, computer processing is now integral to every stage of knowledge generation. It facilitates data access and integration, tracks the origin of data (data provenance), and enables advanced visualization techniques to present findings compellingly. From modeling intricate systems to theorizing based on complex data, technology provides researchers with powerful tools to deepen their understanding and expand the scope of their studies.

Artificial intelligence (AI) has further transformed the research landscape. AI-powered tools can summarize enormous volumes of scientific literature, patents, and databases to uncover trends, patterns, and connections that might otherwise remain hidden. *For example, natural language processing algorithms can rapidly summarize vast amounts of textual data, allowing researchers to focus on critical insights rather than spending hours on manual reviews.*

In applied research, technology opens new avenues for observation and analysis. Eye-tracking technology, *for example, allows researchers to study consumer behaviour by tracking fixation and gaze points as shoppers navigate real or virtual stores.* This provides valuable insights into decision-making processes and marketing strategies.

In upcoming sections, we will explore the pros and cons of technology in research and examine its broader impact on academia. While technology offers tremendous advantages, it also raises important questions about ethical considerations, dependency, and the evolving role of researchers in the age of automation.

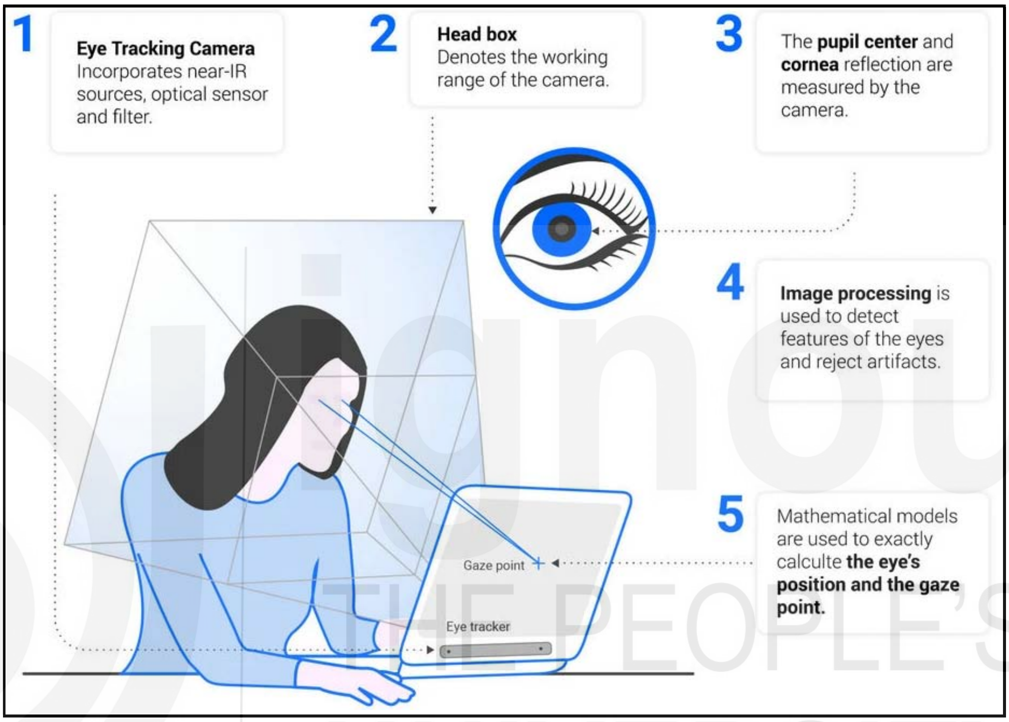


Figure 12.1: Eye Tracking device

12.3.1 Areas Impacted by Technology

Following are the major areas that are impacted by technology:

Technology Impacts on Knowledge: Advancements in information technology have revolutionized the way we access, store, and share knowledge. With digital tools and platforms, vast amounts of information are now readily available, allowing researchers, professionals, and the general public to answer questions, solve problems, and make informed decisions. This access spans various devices, formats, and networks, offering an unprecedented ability to retrieve and process knowledge globally.

Technology Impacts on Academics: In the academic world, technology has transformed teaching, learning, and research. Digital libraries, online courses, and collaboration platforms enable students and educators to access educational materials from anywhere, making learning more flexible and inclusive. Additionally, research databases and citation management tools make it easier for scholars to gather and reference academic work, while advanced software aids in analyzing data, creating simulations, and visualizing complex information.

Technology Impacts on Corporates: Businesses and corporations leverage technology to streamline operations, improve productivity, and innovate. From automation in manufacturing to data analytics for market insights, technology enables companies to make more informed business decisions, enhance efficiency, and compete in the global marketplace. Moreover, digital communication tools facilitate collaboration across geographic boundaries, fostering global partnerships and improving project management.

Technology Impacts on Research and Development: In research and development (R&D), technology has opened up new frontiers. Researchers now use powerful computational tools for simulations, modeling, and data analysis, accelerating discoveries in fields like biotechnology, engineering, and environmental science. AI and machine learning help identify patterns in data that humans might overlook, making the R&D process more efficient and innovative.

Technology Impacts on Governance and Policy: Governments and policymakers rely on technology to create, implement, and monitor public policies. Digital platforms facilitate transparency, citizen engagement, and the efficient delivery of services. Moreover, data analytics help the government to track economic trends, evaluate social programs, and respond to crises like pandemics. Technology also plays a crucial role in cybersecurity, protecting critical infrastructure from cyber threats.

12.4 DATA COLLECTION AND ANALYTICAL TOOLS

Any research process must include data collection and analysis, and technology has completely changed how researchers approach these activities. Researchers may collect data from a variety of sources in the digital era, such as wearable technology, social media platforms, and online questionnaires. In addition to increasing the efficiency of data collecting, these solutions provide users access to previously unreachable respondents, more dynamic datasets.

Online survey tools like Google Forms, SurveyMonkey, and Qualtrics have simplified the process of reaching a wide audience, offering customizable templates, real-time response tracking, and even AI-driven analysis. Social media platforms such as Twitter (now X) and Facebook provide a wealth of unstructured data that can be analyzed using tools like NVivo or Tableau to understand sentiments, trends, and patterns. Wearable devices and sensors, commonly used in healthcare and fitness research, provide real-time data on variables such as heart rate, physical activity, and sleep patterns, adding a layer of granularity to research findings.

In terms of analysis, programs like SPSS, R, Python, Smart PLS, IBM AMOS, and Stata have established themselves as essential tools for quantitative analysis, allowing researchers to carry out tasks ranging from simple descriptive statistics to intricate machine learning models. Making data-driven choices, testing theories, and identifying patterns are all aided by

these technologies. Tools like ATLAS.ti, Orange, NVivo and MAXQDA help with theme analysis and coding for qualitative research, which facilitates the interpretation of textual data, interviews, and narratives.

Additionally, advanced tools powered by artificial intelligence, such as IBM Watson Analytics or Google Cloud AI, offer predictive modeling and natural language processing capabilities, enabling researchers to gain insights from unstructured data like texts, images, or videos. In the below subsection we will discuss some tools related to data collection and analysis.

12.4.1 Google Trends

Google Trends is a powerful tool that allows researchers to analyze the popularity of search queries over time across various regions and languages. By examining trends in search data, researchers can gain insights into public interest and behavior, identify emerging topics, and understand the changing landscape of information. This tool is particularly useful for identifying seasonal patterns, gauging the impact of events on public interest, and exploring correlations between search trends and other data sources, such as social media activity or economic indicators. For academics, Google Trends can inform research questions and hypotheses by highlighting areas that are gaining traction or losing relevance, thus guiding the direction of their studies. By leveraging this tool, researchers can enhance the relevance and impact of their work in a rapidly evolving digital landscape.

Figure 12.2: Google Trends

12.4.2 Typeform

Typeform is a Software-as-a-Service (SaaS) platform designed for creating engaging online forms and conducting dynamic surveys. Its main software creates dynamic forms based on user needs. Typeform's software has been used by Apple Inc, Airbnb, Uber and Nike, Inc. Typeform produces millions of forms every month.

Typeform is an innovative online tool designed for creating engaging surveys, quizzes, and forms that facilitate data collection in a user-friendly manner. Its interactive interface and customizable templates make it easy for researchers to design surveys that capture valuable insights while maintaining participant engagement. Unlike traditional survey tools, Typeform employs a

conversational style, allowing respondents to answer questions one at a time, which can lead to higher completion rates and more thoughtful responses. This tool is particularly useful for gathering qualitative data, conducting market research, and understanding user feedback. Researchers can analyze the collected data in real time, making it easier to draw conclusions and adjust methodologies as needed. By integrating Typeform into their research process, scholars can enhance their data collection strategies and improve the overall quality of their findings.

12.4.3 Statista

Statista is a German online platform that specializes in data gathering and visualization. In addition to publicly available third-party data, Statista also provides exclusive data via the platform, which is collected through its team's surveys and analysis.

Statista is a comprehensive statistics portal that aggregates data from various industries, making it an invaluable resource for researchers seeking reliable and up-to-date information. With access to thousands of statistics, reports, and studies, Statista enables users to easily explore market trends, consumer behavior, and industry insights across diverse sectors. Researchers can utilize this platform to find quantitative data that supports their hypotheses and enriches their analysis. Additionally, Statista provides visualizations such as charts and infographics, which can enhance the presentation of research findings. By leveraging Statista, researchers can save time on data collection while ensuring their work is grounded in credible statistics, ultimately strengthening the quality and impact of their research.

Figure 12.3 Statista

12.4.4 Google Analytics

Google Analytics is one of the most widely used tools for collecting and analyzing data related to website and app performance. It provides researchers, marketers, and businesses with in-depth insights into user behavior, traffic sources, and overall engagement, making it a foundation of digital research and strategy.

Data Collection with Google Analytics

Google Analytics collects data through tracking codes embedded in websites or applications. These tracking codes gather information such as:

- **User Demographics:** Age, gender, and geographical location.
- **Behavioral Data:** Pages visited, time spent on each page, and click-through rates.
- **Traffic Sources:** How users found the website, whether through organic search, direct visits, referrals, or paid campaigns.
- **Devices and Platforms:** Information on the devices, browsers, and operating systems users employ.

This data provides a comprehensive view of how users interact with digital platforms, enabling researchers to identify trends and patterns.

Data Analysis with Google Analytics

Google Analytics offers robust tools for analyzing collected data, including:

- **Audience Analysis:** Understand who your users are and what segments they belong to.
- **Behavior Flow:** Visualize the journey users take across the site, identifying drop-off points and popular content.
- **Conversion Tracking:** Measure how well a website achieves specific goals, such as form submissions, downloads, or purchases.
- **Real-Time Analysis:** Monitor user activity as it happens, providing immediate feedback on the effectiveness of campaigns or updates.

Applications of Google Analytics in Research

For researchers, Google Analytics can be a powerful tool to study user behavior patterns, test hypotheses, and evaluate the impact of changes to digital properties. *For example, academic institutions can analyze how students navigate online learning platforms, while businesses can measure the effectiveness of marketing campaigns*.

12.5 BUSINESS RESEARCH PLATFORMS

Business research tools are mechanisms that help businesses find their target market and determine how the consumers feel about the company's products or services. Business research tools may include data analytics, social media monitoring, competitor intelligence and predictive analytics. There are various kinds of business research which facilitates in gathering sales information and writing a detailed report on marketing and sales. In coming heads we will explore more opportunities and discuss how the essential research tools such as search engines like Google Scholar, JSTOR, and PubMed, reference management software like Zotero, Mendeley, and EndNote, statistical analysis tools like SPSS, R, and Stata, writing tools like Microsoft Word and Grammarly, and data visualization tools like Tableau and Excel will provide more linkages to use business in research.

12.5.1 AnyLogic

AnyLogic is a powerful business research platform that specializes in simulation modeling to support decision-making and strategic planning. It is widely used across industries such as healthcare, logistics, manufacturing, and supply chain management. Any Logic enables researchers and organizations to visualize, analyze, and predict complex system behaviours using advanced simulation techniques.

Key Features of AnyLogic

1. **Multimethod Simulation Modeling:** AnyLogic is unique in offering three types of simulation methods—agent-based modeling (ABM)**,**

discrete event simulation (DES), and system dynamics (SD)—all in a single platform. This flexibility allows researchers to model processes and systems at different levels of detail depending on the research goals.

2. **Dynamic Modeling and Scalability:** The platform can handle large-scale systems with dynamic interactions, making it suitable for exploring scenarios that involve uncertainty, variability, and complex interdependencies. For instance, businesses can simulate supply chain disruptions to identify potential risks and solutions.

3. **Data Integration:** AnyLogic supports integration with various data sources, including Excel, SQL databases, and APIs, allowing users to incorporate real-world data into their simulations. This ensures models are accurate and relevant to the current business environment.

4. **Visualization and Reporting:** The platform provides interactive dashboards and visual outputs, making it easier to communicate findings to stakeholders. Researchers can create animated simulations, graphs, and reports that clearly illustrate results and insights.

Applications in Business Research

AnyLogic is a valuable tool for in following cases:

- **Market Research:** Simulating consumer behavior and market dynamics to predict trends.
- **Supply Chain Optimization:** Modeling logistics networks to improve efficiency and minimize costs.
- **Healthcare Systems:** Analyzing patient flow and resource allocation in hospitals.
- **Manufacturing:** Optimizing production processes and resource planning.

For example: *A logistics company might use AnyLogic to model the impact of delivery delays caused by weather disruptions. By simulating different scenarios, the company can identify optimal routes, resource allocations, and contingency plans to mitigate risks*

12.5.2 R Discovery

R Discovery is a cutting-edge platform designed to simplify the process of academic research by providing researchers with personalized access to the latest scholarly articles and insights. It was developed by cacris communications, a global scientific communications combat, R Discovery leverages artificial intelligence and machine learning to curate and recommend relevant research papers from millions of scholarly sources, saving researchers time and effort.

Key Features of R Discovery

1. **Personalized Recommendations:** R Discovery uses AI-driven algorithms to understand a researcher's area of interest and delivers tailored recommendations. Researchers can create a profile, specify their fields of study, and receive curated content that aligns with their academic focus.

2. **Comprehensive Database Access:** The platform aggregates content from over 90 million research articles, including those from leading journals, open-access repositories, and preprint servers. This ensures users have access to a diverse range of academic literature.

3. **Efficient Literature Reviews:** By offering summarized insights for each article, R Discovery helps researchers quickly grasp the essence of a study. This is particularly useful for conducting literature reviews and identifying key papers in a specific domain.

4. **User-Friendly Interface:** R Discovery's intuitive design allows users to browse, save, and organize articles effortlessly. Researchers can create reading lists, mark articles for later review, and track trends in their field of interest.

5. **Cross-Platform Availability:** The platform is accessible via web and mobile apps, enabling researchers to stay updated on the go.

Figure 12.4 R Discovery

12.5.3 Open Journal Systems (OJS)

Open Journal Systems (OJS) is an open-source platform designed to facilitate the management, publication, and dissemination of scholarly journals. Developed by the Public Knowledge Project (PKP), OJS is widely adopted by academic institutions, publishers, and research organizations to streamline the editorial process and enhance the accessibility of scholarly work.

Key Features of OJS

1. **Editorial Workflow Management:** OJS supports the entire journal publishing workflow, including manuscript submission, peer review, editing, and publication. Its intuitive interface simplifies task assignment, communication, and decision-making for editors, reviewers, and authors.

2. **Customizable and Scalable:** OJS allows journal managers to customize the platform according to their journal's needs. From visual themes to submission guidelines, users can tailor the system to align with their branding and operational requirements.

3. **Open Access Publishing:** Supporting the open-access movement, OJS provides tools for publishing journals in an open-access format, ensuring that research is freely available to the global community without paywalls.

4. **Comprehensive Metadata Management:** The platform enables the inclusion of detailed metadata for each article, making it easier for search engines, indexing services, and readers to discover published content.

5. **Integration with Indexing and Archiving Services:** OJS integrates with services like DOAJ (Directory of Open Access Journals), CrossRef, and PubMed, ensuring widespread visibility and discoverability of journal articles.

6. **Multilingual Support:** OJS supports multiple languages, enabling journals to reach diverse audiences and operate in multilingual environments.

7. **Advanced Reporting and Analytics:** Built-in tools provide insights into article views, downloads, and reader engagement, helping publishers assess the reach and impact of their journals.

Applications in Academic Publishing

OJS is used by universities, research institutions, and independent publishers to manage journals efficiently and promote accessibility. For example:

- Academic institutions can host multiple journals on a single OJS installation, providing a centralized system for various departments.
- Independent publishers can leverage OJS to reduce operational costs and ensure seamless management of their editorial process.

Check Your Progress A:

1) Explain how technology has transformed data collection and analysis in research. Provide examples of tools used for quantitative and qualitative research.

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

2) What are the major areas impacted by technology in research and development? Provide examples to illustrate its role in knowledge, academia, and governance.

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

3) How does Google Analytics aid researchers in understanding user behaviour? Discuss its data collection and analysis capabilities with examples.

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

4) Discuss the role of Artificial Intelligence (AI) in enhancing research processes.

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

12.6 CLOUD COMPUTING

Cloud computing refers to the delivery of computing resources over the internet, often referred to as "the cloud," on demand. These resources include storage, databases, networking, software, analytics, and other services. Rather than relying on local servers or personal devices, cloud computing allows users to access these resources remotely, offering scalability, cost-efficiency, and flexibility.

For researchers, cloud computing has revolutionized data storage and access. It eliminates the need for heavy hardware infrastructure by providing centralized storage solutions, enabling researchers to securely store vast datasets and access them anytime, anywhere. This remote accessibility is particularly valuable for collaborative research, allowing teams across different locations to work together seamlessly on shared projects and datasets.

Cloud computing is an integral part of many tools and platforms used daily. Common examples include:

- **Emails and Calendars**: Platforms like Gmail, Outlook, and Google Calendar are cloud-based, allowing access to information across multiple devices.
- **File Storage Services**: Tools such as Google Drive, Dropbox, and OneDrive store documents, presentations, and multimedia securely in the cloud, enabling instant sharing and collaboration.
- **Research Tools:** Platforms like AWS (Amazon Web Services), Microsoft Azure, and Google Cloud Platform offer advanced cloud computing services tailored for researchers, including high-performance computing for simulations and analyses.

- **Collaboration Tools:** Applications like Google Workspace, Slack, and Trello facilitate real-time communication and project management through cloud-based infrastructure.

There are various other examples given in Table 12.1

Table 12.1 Different Types of Cloud computing service

Sl. No.	Title	Description	Brand Logo
1.	**Skype**	Skype is a proprietary telecommunications application operated by Skype Technologies, a division of Microsoft, best known for IP-based video telephony, videoconferencing and voice calls. It also has instant messaging, file transfer, debit-based calls to landline and mobile telephones, and other features.	
2.	**Whats App**	WhatsApp is an instant messaging and voice-over-IP service owned by technology conglomerate Meta. It allows users to send text, voice messages and video messages, make voice and video calls, and share images, documents, user locations, and other content	
3.	**Google Cloud**	Google Cloud is a suite of cloud computing services that runs on the same infrastructure that Google uses internally for their own consumer products, such as Google Search, Gmail, and YouTube. The list of available Google Cloud services is long—and it keeps growing.	
4	**Amazon Web Services (AWS)**	Amazon Web Services, Inc. is a subsidiary of Amazon that provides on-demand cloud computing platforms and APIs to individuals, companies, and governments, on a metered, pay-as-you-go basis. Clients will often use this in combination with autoscaling	

If we talk in a present scenario the Cloud computing is the on-demand access of computing resources—physical servers or virtual servers, data storage, networking capabilities, application development tools, software, AI-powered analytic tools and more—over the internet with pay-per-use pricing

There are four different types of cloud computing: public, private, hybrid, and community. They vary in their levels of accessibility, security risks, costs, and other factors, and each one carries its own advantages and disadvantages of cloud computing from Researcher point of view:

Advantages of cloud computing for Researchers:

1. **Scalability**: Cloud platforms can adjust resources based on research needs, whether it's expanding storage for larger datasets or increasing computational power for complex simulations.

2. **Cost-Effectiveness**: Pay-as-you-go pricing models reduce upfront costs, making high-end computing resources accessible to individual researchers and institutions alike.

3. **Data Security**: Reputable cloud services provide robust security measures, including encryption and regular backups, to protect sensitive research data.

4. **Collaboration**: Researchers can work in real-time with peers globally, leveraging cloud services for document sharing, video conferencing, and synchronized editing.

5. **Data Loss Prevention**: Unlike traditional local storage, cloud computing significantly reduces the risk of data loss. Automatic backups, redundancy, and disaster recovery protocols ensure that data remains intact and accessible even in the event of hardware failure, theft, or natural disasters.

Limitations of Cloud Computing for Researchers:

1. **Dependence on Internet Connectivity:** Cloud computing relies heavily on stable internet access. Any disruption in connectivity can hinder access to resources, particularly in remote areas with poor network infrastructure.

2. **Cost Over Time:** While cloud computing reduces upfront costs, the pay-as-you-go model can become expensive over time, especially for large-scale projects requiring extensive storage or computing power.

3. **Data Privacy Concerns:** Storing sensitive research data on third-party servers raises concerns about data privacy and compliance with regulations like GDPR or HIPAA. Mismanagement or breaches by cloud service providers can compromise confidential information.

4. **Limited Control:** Researchers may have limited control over the infrastructure and services provided by cloud vendors. Any changes or downtimes on the provider's end can impact ongoing research projects.

5. **Potential for Vendor Lock-In:** Migrating from one cloud service provider to another can be challenging due to compatibility issues and the complexity of data transfer processes, which may restrict flexibility in choosing or switching providers.

12.7 VIRTUAL AND AUGMENTED REALITY

Virtual Reality (VR) and Augmented Reality (AR) are cutting-edge technologies that are transforming the way research is conducted across various fields. VR creates a fully immersive digital environment, allowing researchers to simulate complex scenarios, conduct experiments, and analyze human behavior in a controlled yet realistic setting. On the other hand, AR overlays digital information onto the physical world, making it particularly useful for enhancing real-time data collection and visualization. These technologies are being employed in fields such as healthcare, education, and social sciences to simulate environments that are difficult or costly to recreate in the physical world. For researchers, VR and AR offer new ways to visualize data, collaborate remotely, and engage participants in ways that were previously unimaginable. By integrating these technologies, researchers can push the boundaries of their studies, enhance interactivity, and generate deeper insights into their areas of inquiry.

Figure 12.5 Augmented and Virtual Reality

Application of Virtual and Augmented Reality in different are of research:

- **Behavioural and Psychological Studies:** VR provides a controlled environment to simulate scenarios for studying human responses, decision-making, and emotional reactions. For instance, researchers can create virtual environments to study phobias, PTSD treatments, or social behaviors.

- **Healthcare and Medical Research:** In healthcare, AR helps overlay body details on patients during surgeries, while VR enables medical students and professionals to practice complex procedures in a risk-free virtual environment.

- **Design and Prototyping:** VR and AR allow researchers in engineering, manufacturing, and architecture to visualize, design, and test prototypes in a virtual space, reducing costs and errors before physical production.

- **Education and Training:** AR applications bring textbooks to life with interactive 3D models, while VR offers immersive environments for experiential learning and skill development in fields like aviation, medicine, and construction.

- **Consumer Behavior and Marketing:** VR and AR are employed to study consumer engagement by creating immersive shopping environments or testing product placements in real-world contexts.

Benefits for Researchers

These technologies enhance data collection, improve precision, and enable cost-effective simulations. *For example, VR allows researchers to record interactions in virtual worlds, while AR facilitates real-time visualizations of complex systems.*

Challenges for Researchers

Despite their immense potential, barriers such as high costs, limited accessibility, and the need for technical expertise can hinder widespread adoption.

12.8 COLLABORATION TOOLS

Collaboration tools play a crucial role in modern research by enabling interdisciplinary and cross-institutional cooperation. These tools support researchers in achieving common goals, such as developing innovative solutions, conducting experiments, analysing data, and publishing results. By facilitating communication, task management, and resource sharing, collaboration tools streamline the research process, improve efficiency, and foster global connectivity.

Types of Collaboration Styles in Research

- **Communication-Oriented Collaboration:** This style emphasizes seamless communication among research team members, particularly important for discussions, brainstorming sessions, and feedback. Tools like video conferencing and instant messaging allow researchers from different time zones to connect in real-time.

- **Task-Oriented Collaboration:** Researchers often collaborate on complex projects with well-defined objectives, timelines, and deliverables. Task-oriented tools, such as project management platforms, help allocate roles, track progress, and maintain accountability.

- **Network-Oriented Collaboration:** Research often involves building networks with other scholars, institutions, and industry partners. Tools that facilitate collaborative writing, data sharing, and networking are essential to expand knowledge and resources across disciplines.
- **Community-Oriented Collaboration:** Research communities thrive on sharing insights, resources, and experiences. Platforms that support shared knowledge, forums, and repositories enable researchers to create a collaborative ecosystem.

Some examples of Collaboration Tools:

1. **Zoom Workplace**: An AI-powered, open collaboration platform, Zoom Workplace streamlines communication, increases employee engagement, and optimizes in-person and virtual interactions. It integrates with third-party apps, offering seamless workflows for modern teams.
2. **Microsoft Teams**: A robust platform combining chat, video calls, file sharing, and integration with Microsoft Office tools, enabling real-time collaboration within and across organizations.
3. **Overleaf**: An online LaTeX editor that enables researchers to collaboratively write and edit academic papers in real-time, ensuring proper formatting for publications.
4. **Mendeley**: A reference management tool that allows teams to organize citations, share reading lists, and annotate research articles.
5. **Slack**: Widely used for communication, Slack enables researchers to create channels for project discussions, share files, and integrate with other tools for seamless collaboration.
6. **Google Workspace**: Tools like Google Docs, Sheets, and Drive allow multiple researchers to co-author documents, share datasets, and store resources securely in the cloud.

Following are the benefits of Collaboration Tools in Research

- **Global Collaboration:** Researchers from different parts of the world can work together seamlessly, regardless of geographical barriers.
- **Resource Sharing:** Tools provide centralized access to shared datasets, articles, and experiment results.
- **Efficiency:** Simplified communication and task management help researchers focus on their core activities.
- **Knowledge Exchange:** Researchers can engage in discussions, share expertise, and build networks to enhance the quality of their work.

Check Your Progress B:

1) Examine the role of cloud computing in collaborative and data-driven research.

   …………………………………………………………………………….

   …………………………………………………………………………….

2) Give some examples of Collaboration tools?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

3) Compare Virtual Reality (VR) and Augmented Reality (AR) in research applications.

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

4) Choose the most appropriate answer

i) What is one of the key advantages of cloud computing in research?

a) High upfront cost

b) Scalability

c) Dependence on local servers

d) Limited accessibility

ii) Which of the following is an example of a cloud-based file storage service?

a) NVivo

b) Google Drive

c) IBM Watson Analytics

d) SPSS

iii) Which collaboration tool is known for enabling real-time co-authoring of documents?

a) Slack

b) Overleaf

c) Mendeley

d) Google Workspace

iv) What is a key application of Virtual Reality (VR) in research?

a) Simulating complex scenarios

b) Real-time collaboration

c) Cloud storage

d) Generating statistical reports

v) Which platform supports both communication and project management in research collaborations?

a) Overleaf

b) Slack

c) IBM Watson Analytics

d) Typeform

12.9 ROBOTICS AND AUTOMATION IN RESEARCH

Robotics and automation have revolutionized research by enabling tasks that require precision, efficiency, and consistency to be performed at an unprecedented scale. By integrating intelligent machines and automated systems, researchers can conduct experiments, analyze data, and manage repetitive tasks more effectively. This not only reduces human error but also frees up researchers to focus on higher-order cognitive tasks, such as interpreting results and developing new hypotheses.

Applications of Robotics in Research

1. **Laboratory Automation:** Robotics is widely used in automating routine laboratory tasks such as pipetting, mixing solutions, and handling samples. *For example, liquid-handling robots automate the preparation of samples for high-throughput screening in pharmaceutical research, significantly reducing manual labour and increasing efficiency.*

2. **Field Research:** In environmental studies, robots like underwater drones and aerial drones are used to collect data from remote or hazardous locations. For instance, autonomous underwater vehicles (AUVs) can map ocean floors and monitor marine ecosystems, providing data that would otherwise be difficult or dangerous to obtain.

3. **Medical Research and Healthcare:** Robotic systems such as Da Vinci Surgical System are used in minimally invasive surgeries, while automation assists in developing personalized medicine by analyzing genetic data with precision. Robotic arms are also employed in drug discovery to perform repetitive experiments efficiently.

4. **Artificial Intelligence in Robotics:** Combining robotics with AI allows for adaptive learning and decision-making in tasks like sample sorting, data collection, and experiment monitoring. AI-powered robots are used in agricultural research to study crop patterns and monitor plant health.

Examples of Automation in Research

1. **Data Analysis:** Automated software tools like Python scripts or R packages are commonly used to process large datasets, identify patterns, and generate predictive models, saving researchers hours of manual computation.

2. **Literature Review Automation:** Tools such as Elicit.AI and Sci Space automate the process of finding and organizing relevant academic literature, streamlining the initial stages of research.

3. **Automation in Manufacturing Research:** Automated systems simulate real-world manufacturing processes in controlled environments, enabling engineers to test the efficiency and safety of new technologies. *For example, robotic assembly lines are used to research industrial efficiency and process optimization.*

Benefits of Robotics and Automation in Research:

1. **Precision and Consistency**: Robots can perform tasks with exacting accuracy, reducing the variability in experimental outcomes.

2. **Efficiency**: Automation speeds up repetitive processes, allowing researchers to complete large-scale projects in less time.

3. **Safety:** Robots can operate in hazardous environments, reducing risks to researchers.

4. **Scalability**: Automated systems can handle larger datasets and more experiments simultaneously, making them ideal for big data and high-throughput research.

Despite their benefits, robotics and automation come with challenges, such as high initial costs, technical complexity, and the need for specialized training. Additionally, ensuring ethical use and addressing concerns about the replacement of human roles in research remain critical.

12.10 AI IN RESEARCH

Artificial Intelligence (AI) has emerged as a transformative force in research, enabling researchers to handle complex data, automate repetitive tasks, and derive insights that would be difficult or impossible to uncover through traditional methods. AI encompasses machine learning, natural language processing (NLP), computer vision, and other advanced techniques that mimic human intelligence. In research, AI is not just a tool; it is a collaborator that enhances every stage of the research process, from hypothesis generation to data analysis and beyond.

Artificial intelligence (AI) is a key tool in the scientific process, helping researchers in many ways, including:

Data Analysis and Pattern Recognition: AI can process and analyze large datasets quickly, identifying trends and correlations. *For example, machine learning models are used in genomics to analyze DNA sequences and uncover patterns related to diseases.*

Natural Language Processing for Literature Reviews: NLP-powered tools like Elicit.AI and Semantic Scholar assist researchers by summarizing academic papers, identifying key themes, and generating connections across multiple sources of literature.

Predictive Modelling: AI is widely used to create predictive models in fields like climate science, finance, and epidemiology. *For example, AI models can predict disease outbreaks by analyzing environmental and demographic data.*

Automated Experimentation: AI systems can design, conduct, and optimize experiments autonomously. For instance, in drug discovery, AI algorithms simulate chemical interactions to identify promising drug candidates faster than traditional methods.

Visual Data Analysis: Computer vision, a subset of AI, analyzes images and videos. In medical research, AI-powered systems like DeepMind's, Alpha Fold predict protein structures with high accuracy, accelerating research in biology and medicine.

Personalized Research Recommendations: AI tools like R Discovery provide researchers with curated articles and insights tailored to their specific areas of interest, saving time and effort.

Following are the benefits of AI in Research:

- **Efficiency:** AI reduces the time required for data processing, literature review, and repetitive tasks.
- **Improved Insights:** AI helps identify patterns and generate insights from complex datasets that are difficult for humans to analyze manually.
- **Enhanced Collaboration:** AI tools like Research Rabbit foster interdisciplinary studies by linking related research across diverse fields.
- **Scalability:** AI can process vast amounts of data, making it indispensable for big data research.

Following are the challenges of AI in Research

- **Data Quality and Bias:** AI outcomes depend heavily on the quality and representativeness of input data. Poor data can lead to flawed conclusions.
- **Ethical Concerns:** Issues like data privacy, algorithmic bias, and the potential misuse of AI tools must be carefully managed.
- **Accessibility:** Advanced AI tools may require significant computational resources, making them less accessible to underfunded researchers or institutions.
- **Skill Gap:** Effective use of AI requires technical expertise, which may not be readily available to all researchers.

12.10.1 Generative AI

Generative AI is becoming an essential tool in research, offering innovative ways to assist researchers at different stages of their work. Unlike traditional AI, which follows strict instructions, generative AI can create new content, such as text, images, and even music, based on patterns and data it has learned. This makes it easier for researchers to come up with new ideas and solutions, all while saving time and boosting creativity.

How Generative AI Helps in Research:

1. **Automating Experiment Design:** Generative AI can help design experiments by suggesting methods, generating ideas, and even predicting possible outcomes. This saves researchers time by reducing the effort needed to plan studies and allows more focus on analyzing data.

2. **Brainstorming Partner:** Tools like **ChatGPT** are great for brainstorming. Researchers can use them to generate ideas and explore different approaches to their research problems. This is especially helpful in the early stages of planning a project when a lot of creativity is needed.

3. **Generating Content:** One of the main uses of generative AI is in creating written content. For example, ChatGPT can help researchers draft parts of papers, reports, or summaries, making the writing process faster and easier.

4. **Improving Research Efficiency:** Generative AI can speed up tasks that would normally take a lot of time, such as writing drafts or analyzing large amounts of information. It also helps avoid mistakes and ensures that researchers stay productive by handling repetitive tasks.

12.10.2 Popular AI Tools in Research

Many AI tools are now available to help researchers perform their work more effectively refer Figure 12.8. Here are a few examples:

Elicit.AI: An AI tool designed to assist with literature reviews by summarizing key findings, identifying related research, and organizing information efficiently.

Research Rabbit: Known as a "Spotify for Research," this platform helps researchers discover and explore papers in a visual and interactive manner. It maps relationships between studies, enabling quick identification of influential works and emerging trends.

Consensus: An AI tool specifically designed for answering research questions. It synthesizes evidence from academic papers to provide direct and reliable insights, making it ideal for hypothesis validation.

DeepMind's Alpha Fold: Revolutionizing biological research, Alpha Fold predicts protein structures with high accuracy, accelerating drug discovery and genomics research.

IBM Watson Analytics: Offers natural language processing and predictive analytics for complex data analysis across various research fields.

Connected Papers: Visualizes relationships between academic papers, making it easier for researchers to understand how ideas and studies are interconnected.

SciSpace: It simplifies writing and formatting academic papers by integrating AI to streamline the publication process.

H2O.ai: An open-source AI platform for predictive analytics and data visualization, ideal for researchers handling large datasets.

Julius.AI: Julius is an AI Data Scientist that can analyze and visualize massive datasets, perform complex analysis like forecasting and regression, and even train Machine Learning models.

QuillBot: It was developed in 2017, is an AI-powered tool designed to paraphrase, summarize, and refine text, making it an invaluable resource for researchers, students, and professionals. It is particularly useful for rephrasing complex academic content, ensuring readability, and avoiding plagiarism. QuillBot's summarizer condenses lengthy articles into concise summaries, allowing researchers to quickly grasp key ideas without extensive reading. For non-native English speakers, it provides language assistance, improving grammar, sentence structure, and word choice to meet academic standards. While it generate citations, QuillBot also complements tools like Mendeley and Zotero by effectively paraphrasing cited content for literature reviews. Additionally, it helps align content with the formal tone and structure required in academic writing. This versatile tool streamlines the writing process, saving time and enhancing the clarity and quality of research outputs.

BIT AI : Bit AI is a smart document collaboration tool designed for teams and individuals to create, manage, and share documents seamlessly. It enables real-time collaboration, allowing multiple users to work on the same document simultaneously, which makes teamwork faster and more efficient. With features like media integration, users can enrich their documents by embedding videos, links, charts, and files directly, ensuring that all resources are centralized and easily accessible. Bit AI also provides pre-made templates, simplifying the process of creating reports, research papers, presentations, and other content. For instance, a research team can use Bit AI to collaboratively draft a paper, track edits, and incorporate resources like videos or datasets into the document in a user-friendly manner, streamlining the entire research workflow.

Grammarly: It is an AI-powered writing assistant that enhances the clarity, accuracy, and quality of academic writing. It identifies grammatical errors, corrects spelling and punctuation, and suggests improvements for clarity, conciseness, and tone, ensuring research documents maintain a formal and professional style. Grammarly is particularly useful for non-native English speakers, providing guidance to meet academic standards. With features like integration into platforms such as Microsoft Word and Google Docs, and a plagiarism detection option in its premium version, Grammarly streamlines editing and ensures originality. While it excels in general writing, users may need to manually adjust domain-specific language to maintain context. Overall, Grammarly is a valuable tool for researchers seeking to produce polished, high-quality work efficiently.

Figure 12.6 Grammarly

ChatGPT: It is developed by OpenAI. It is an advanced AI language model that assists researchers across various stages of their work. It is particularly useful for brainstorming ideas, drafting content, simplifying complex concepts, and refining academic writing. Researchers can leverage ChatGPT to generate initial drafts for papers, proposals, or reports, as well as seek guidance on coding, statistical concepts, or data analysis approaches. Its ability to break down intricate theories into simpler explanations makes it valuable for teaching and communicating research findings. Additionally, ChatGPT can help enhance clarity and tone in written work, making revisions more efficient. While it offers significant time-saving benefits and versatility, its outputs must be validated for accuracy, especially for technical or specialized content. Despite its limitations, ChatGPT serves as a powerful tool for enhancing productivity and streamlining the research process.

Figure 12.7: Chat GPT

Scite: It is an innovative tool designed to provide both quantitative and qualitative insights into how scientific publications cite each other. While traditional citation metrics focus solely on the number of times a paper is cited, Scite goes beyond this by analyzing the context and purpose of each citation. Using access to full-text articles and deep learning models, Scite identifies the textual context of citations, showing where and why a citation was made.

For any given publication, Scite can:

- Count how many times it was cited.
- Display the text from citing papers where the citation occurred, offering insight into the context.
- Classify citations as supporting, contrasting, or simply mentioning the cited claims, helping researchers evaluate the influence and credibility of a study.

This qualitative approach enables researchers to understand not just the volume of citations but the intent behind them, making Scite a valuable tool for evaluating the impact and reliability of scientific work. By providing nuanced citation analysis, Scite enhances the research process and fosters a deeper understanding of the academic landscape.

AI Software	**Primary Use**
Elicit.AI	Literature Review
Research Rabbit	Literature Review
Consensus	Literature Review
DeepMind's AlphaFold	Data Analysis
IBM Watson Analytics	Data Analysis
Connected Papers	Literature Review
SciSpace	Literature Review
H2O.ai	Data Analysis
Julius.AI	Data Analysis
QuillBot	Literature Review
Bit AI	Literature Review
Grammarly	Literature Review
ChatGPT	Data Analysis and Literature Review
Scite	Literature Review

Figure 12.8 Popular AI Software and their Applications in Research

AI research involves designing, developing, and analyzing algorithms, models, and systems that enable machines to perform tasks that typically require human intelligence

12.11 FREE AND OPEN-SOURCE SOFTWARE

Free and open-source software is software that is available under a license that grants the right to use, modify, and distribute the software, modified or not, to everyone free of charge. The public availability of the source code is, therefore, a necessary but not sufficient condition. Free and open-source software (FOSS) is an umbrella term for software that is simultaneously considered both free software and open-source software. The precise definition of the terms "free software" and "open-source software" applies them to any software distributed under terms that allow users to use, modify, and redistribute said software in any manner they see fit, without requiring that they pay the author(s) of the software a royalty or fee for engaging in the listed activities. Free and open-source software (FOSS) is software that is available under a license that grants the right to use, modify, and distribute the software, modified or not, to everyone free of charge refer table 12.2 to know the difference between open source software and closed source software. The public availability of the source code is, therefore, a necessary but not sufficient condition. FOSS is an inclusive umbrella term for free software and open-source software.

FOSS is in contrast to proprietary software, where the software is under restrictive copyright or licensing and the source code is hidden from the users. Refer figure 12.9 & 12.10 for different types of software & open source software and proprietary software respective.

	Free software	Open-source software	Freeware	Public-domain software
Definition	"FREE" is a matter of liberty, not price	"OPEN" doesn't just mean access to the source code	"FREE" refers to price, while freedom of the use is restricted by creator	"PUBLIC DOMAIN" belongs to the public as a whole
Ground philosophy	Social movement	Development methodology	Marketing goals	Copyright disclamation
Ground rules	Four Freedoms https://www.gnu.org/philosophy/free-sw.html	Open Software initiative https://opensource.org/osd		Creative Common Organization https://creativecommons.org
Free of charge	Not necessary	Not necessary	✓YES	✓YES
Covered by copyright law	✓YES	✓YES	✓YES	✗ NO
Examples	Linux, ubuntu	MySQL, Apache	Skype, Adobe Acrobat	SQLite

Figure 12.9 Different Types of Software

Table 12.2: Difference between Open Source Software and Closed Source Software

Open source software	Closed source software
Source code is open to all	Source code is closed/protected– Only those who created it can access it
Open source software license promotes collaboration and sharing	Proprietary software license curbs rights
Less costly	High-priced
Less restriction on usability and modification of software.	More restrictions on usability and modification of software.
Big and active community enabling quick development and easy fixes	Development and fixes depend on the discretion of creators.
Support is through forums, informative blogs, and hiring experts	Dedicated support
Immense flexibility as you can add features, make changes, etc.	Limited flexibility (only as proposed by its creators)
Developers are ready to offer improvements hoping to get recognition.	Need to hire developers to integrate improvements.
Can be easily installed into the computer	Needs valid license before installation
Fails and fixes fast	Failure is out of the question
No one is accountable for any failures	Responsibility for failure clearly rests on the vendor

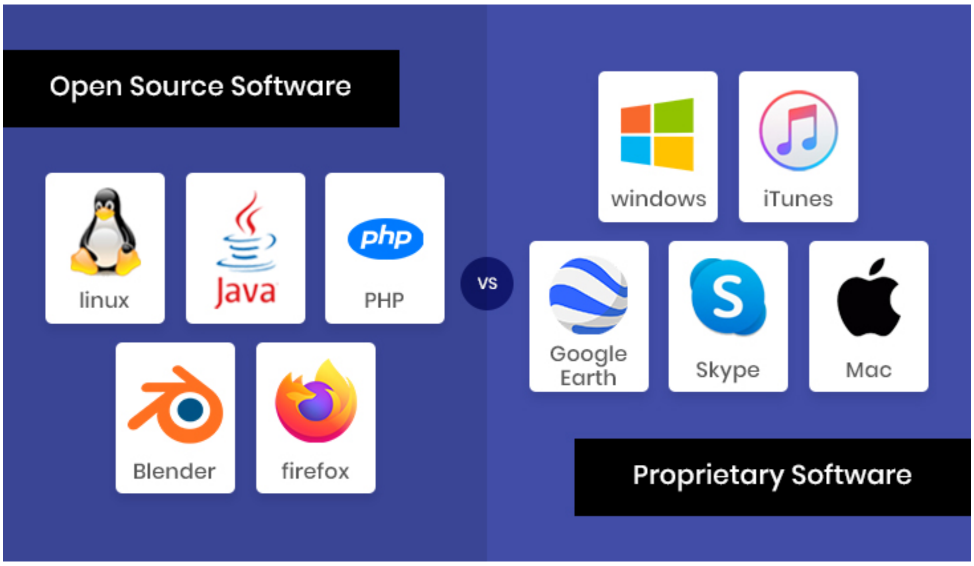


Figure 12.10: Open source Software and Proprietary Software

12.12 WORD PROCESSING SYSTEM (WPS)

WPS Office (Writer, Presentation, Spreadsheet) is a comprehensive productivity software suite similar to Microsoft Office, designed for creating documents, presentations, and spreadsheets. Known for its user-friendly interface and cost-effectiveness, WPS is an excellent tool for researchers working across various platforms.

Key Features of WPS Office

1. **WPS Writer:** A word processing tool similar to Microsoft Word, ideal for drafting and editing research papers, theses, and reports.
2. **WPS Presentation:** A presentation tool akin to Microsoft PowerPoint, used for creating visually appealing slides to showcase research findings.
3. **WPS Spreadsheet:** A spreadsheet application comparable to Microsoft Excel, suitable for data analysis, creating charts, and performing calculations.
4. **Cross-Platform Accessibility:** Available on Windows, macOS, Android, and iOS, ensuring seamless work across devices.
5. **Cloud Storage Integration:** Offers cloud storage, allowing researchers to save documents online and access them from anywhere, facilitating collaboration and flexibility.

Applications in Research

- **Document Creation:** Researchers can use **WPS Writer** to draft, format, and edit academic documents, ensuring professional presentation and compatibility with standard formats.
- **Data Analysis: WPS Spreadsheet** is ideal for organizing and analyzing research data, performing statistical calculations, and generating graphs.
- **Presentations: WPS Presentation** helps in creating polished slides for conferences, seminars, or group discussions, making it easier to communicate complex findings.

WPS Office is a versatile and accessible suite for researchers, combining essential tools for writing, data analysis, and presentations in one package. Its cross-platform support and cloud integration make it a reliable choice for researchers seeking productivity and convenience in their academic and professional endeavors.

12.13 LET US SUM UP

This unit talks about how technology has transformed research processes, making them more efficient, precise, and innovative. From hypothesis formulation to presenting findings, technology plays a pivotal role. Tools like spreadsheets, data visualization platforms, and AI-powered systems enhance data collection, analysis, and collaboration. Artificial Intelligence (AI), in particular, revolutionizes research by automating repetitive tasks, identifying intricate patterns, and enabling predictive modeling. This has significantly impacted diverse fields such as genomics, behavioral studies, and social sciences.

The integration of technology has reshaped several key areas. Knowledge access has become global, with digital tools and platforms breaking geographical barriers. In academia, resources like digital libraries and specialized software enhance teaching, learning, and research. Businesses utilize technology for data-driven decision-making, innovation, and streamlining operations. In research and development, computational tools and AI accelerate discoveries in management, biotechnology, engineering, and environmental science.

For data collection and analysis, tools like Google Analytics, SPSS, and Python facilitate quantitative studies, while NVivo and MAXQDA support qualitative analysis. Platforms like Google Trends, Type form, and Statista offer dynamic insights into trends and behaviour. Advanced technologies like cloud computing enable scalable, secure, and collaborative research. Virtual and augmented reality transform data visualization and experimentation, creating immersive environments. Collaboration tools such as Google Workspace, Slack, and Mendeley foster teamwork and resource sharing, making research more efficient and interconnected.

Despite these advancements, challenges such as high costs, skill gaps, data privacy concerns, and over-reliance on technology remain a concern. However, by leveraging these tools effectively, researchers can overcome traditional barriers, enhance the quality of their work, and foster innovation across disciplines. Technology thus serves as a catalyst, empowering researchers to push the boundaries of knowledge and discovery.

12.14 KEYWORDS

Artificial Intelligence (AI): A field of computer science focused on creating systems that can perform tasks typically requiring human intelligence, such as pattern recognition, decision-making, and language processing.

Augmented Reality (AR): A technology that overlays digital information (e.g., images, sounds) onto the physical world, enhancing real-time data collection and visualization.

Cloud Computing: The delivery of computing services, including storage, processing, and networking, over the internet, enabling scalable and remote access to resources.

Collaboration Tools: Software platforms that facilitate teamwork, communication, and resource sharing across geographic and organizational boundaries.

Data Analysis: The process of examining, cleaning, transforming, and interpreting data to discover useful information and support decision-making.

Digital Libraries: Online repositories providing access to a vast collection of academic and research resources, breaking barriers of time and geography.

Google Analytics: A web analytics tool used to track and report website traffic, providing insights into user behavior, demographics, and engagement.

Knowledge Access: The ability to retrieve, share, and process information using digital tools, enhancing problem-solving and informed decision-making.

Machine Learning: A subset of AI involving algorithms that improve automatically through experience and data, used for predictive modeling and trend analysis.

Natural Language Processing (NLP): An AI subfield focused on enabling machines to understand, interpret, and respond to human language.

Qualitative Analysis: The examination of non-numeric data like texts, interviews, or images to identify themes and patterns.

Quantitative Analysis: The use of numerical data and statistical methods to analyze and interpret measurable variables.

Robotics: The design and use of robots to perform tasks with precision and efficiency, often used in research for automation and data collection.

Statistical Software: Tools like SPSS and R used for performing statistical analysis, hypothesis testing, and data visualization.

Virtual Reality (VR): An immersive technology that creates a simulated digital environment for controlled experiments and data visualization.

12.15 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress B

4 (i) b (ii) b (iii) d (iv) a (v) b

12.16 TERMINAL QUESTIONS

1) How has Artificial Intelligence (AI) transformed the research process? Provide examples of tools and their applications.
2) Explain the impact of technology on data collection and analysis in research.
3) What are the advantages of using free and open-source software in research?
4) What are the ethical considerations associated with using AI in research?
5) What are the major areas impacted by technology in research? Explain with examples.
6) Examine the role of cloud computing in collaborative and data-driven research.
7) Discuss the significance of collaboration tools in global research projects.

12.17 FURTHER READINGS

1. Fry, L. W. (1982). Technology-structure research: Three critical issues. *Academy of Management Journal*, *25*(3), 532-552.
2. Orlikowski, W. J., & Barley, S. R. (2001). Technology and institutions: What can research on information technology and research on organizations learn from each other?. *MIS quarterly*, 145-165.
3. Venkatesh, V., Davis, F., & Morris, M. G. (2007). Dead or alive? The development, trajectory and future of technology adoption research. *The Development, Trajectory and Future of Technology Adoption Research (April 27, 2007). Venkatesh, V., Davis, FD, and Morris, MG "Dead or Alive*, 267-286.

Note

These questions are helpful to understand this unit. Do efforts for writing the answer of these questions but do not send your answer to university. It is only for your practice.

UNIT 13 SOFTWARE APPLICATION IN RESEARCH

Structure

13.0 OBJECTIVES

After studying this unit, you will be able to:

- Understand the role of various software applications in business research;
- Understand the research's different software types and appropriate applications;
- Comprehend the application of specific software, like SPSS, Microsoft Office, R, Tableau, and others;
- Know the tools for collection of data, citation, and reference management; and
- Learn about the ethical considerations of using research software and tools.

13.1 INTRODUCTION

In the preceding unit, we have discussed the usage of technology in research. Now, in this unit we are going to study about the various software applications and their usage in research. Software applications are essential elements in the field of research. It helps researchers in every stages of research. Current trends suggest that researchers can utilize the emerging technologies like artificial intelligence, machine learning, blockchain, and augmented reality for determining target population, samples, creating questionnaires, and collecting data. A variety of software are available that assist in questionnaire design, sample size calculation, and further tasks.

Software is the core component of many instruments used in research. The term "instrument" needs to be understood in a broad sense, encompassing various sorts of both physical and virtual instruments across diverse research disciplines. In the social sciences, survey software serves as an instrument, with a component being user-facing software utilized for data collection (applications, websites, etc.).The role of research software in our instruments encompasses data acquisition, techniques for monitoring or exporting experimental data, as well as data cleaning and processing.

In business research, the term "instrument" includes both physicals digital tools used for gathering and analyzing data. In areas like marketing and consumer behaviour, tools such as survey software and sentiment analysis progress help researchers collect the data through online surveys, website tracking, and social media analysis. Research software plays an important role in collecting data, tracking marketing trends in real-time, exporting data, and cleaning and processing information to generate useful insights for decision making.

This unit covers the variety of software tools available for research, which emphasize their characteristics, benefits, and optimal application techniques. We will discuss statistical software, databases, reference management software, and citation metrics. By understanding these tools and their utilisations, researchers may improve their workflows, enhance the quality of their research, and acquire extensive knowledge from their data.

As research progresses, expertise in these software tools is essential for researchers aiming at producing genuine well-organized, and impactful work.

13.2 SOFTWARE APPLICATIONS IN RESEARCH

Software applications are very important for recent time researchers because they make data more efficient, accurate, and easy to access. These tools help researchers at various stages in the research process, such as when they are collecting data, analyzing it, writing reports, and working together.

Some poll tools, like Qualtrics and Google Forms, are used to collect data. They make it easy to get information from participants and often come with built-in tools for organizing and visualizing the data. Software like NVivo and ATLAS.ti can help you organize, code, and look at big amounts of textual or multimedia data for qualitative research.

Statistical programs like SPSS, R, and Python are used to do complex data analysis in quantitative research. These programs help with things like regression models, hypothesis testing, and trend analysis. These tools make math faster, which means fewer mistakes made by people. This makes the results more reliable. Platforms like MATLAB and Stata are used a lot for advanced modeling and simulations by researchers who work with big datasets.

Reference management tools like EndNote, Mendeley, and Zotero make it easy to organize sources and make bibliographies, which makes literature reviews and academic writing more efficient. Overall, using software applications in research makes the research process better by allowing for more thorough analysis, encouraging collaboration, and helping to make sure that results are shared correctly and on time.

13.2.1 System Software

System software is a type of computer program that manages a computer's resources and runs its hardware and applications. It's typically bundled with a computer's operating system and provides a platform for running application software. It includes operating systems like Linx, macOS, Windows, which facilitates the usage of other research software.It is intended to control the hardware resources of a computer system to provide applicable usefulness to the user.

In research systems softwares can be used in a variety of ways, which includes; studying or developing softwares, creating or analyzing data organization, presenting information to a user, research information management etc.,. The purpose of research software as a component in research instruments can include acquisition, methods to stream or upload experimental data, data cleaning, and processing. Or more generally, research software components organize, serve, and provide access to data.

13.2.2 Application Software

Application software generally known as App, are often called productivity programs or end-user programs because they enable the user to complete

tasks, such as creating documents, spreadsheets, databases and publications, doing online research, sending email, designing graphics, running businesses, and even playing games.

Application software is a powerful tool for research, and that can be very helpful in improving the accuracy, efficiency, and effectiveness of the research process. Some of the application software that are mostly used in research includes the following:

- **Microsoft Word:** It is a prevalent word processing application utilized for composing essays, research papers, dissertations, and various other documents.
- **Spreadsheet Software:** It is a tool utilized for computations, extensively employed in data analysis in both research and industry.
- **Web Browser:** It facilitates user access to web pages and enables many research activities.
- **Simulation software:** It is mostly utilized by engineers to simulate real-world phenomena prior to product manufacturing.
- **Project Management Software**: This software is utilized for project planning, scheduling, resource allocation, and other similar purposes.

Researchers need to carefully evaluate the pros and cons of using application software, including the time it takes to learn, concerns about compatibility, data security, cost, and ethical issues.

13.2.3 Data Analytical Software

Data analytics software, commonly referred to as business analytics software, uses statistical techniques to organize, categorize, and analyze business data from multiple sources while presenting it in an accessible way to facilitate informed decision-making. Data analytics transforms unprocessed data into useful information. It encompasses different tools, and techniques designed to identify trends and manage issues through data analysis. Data analytics may affect business operations, enhance decision-making, and promote business growth. Data analytics is frequently mistaken for data analysis. Although these concepts are interconnected, they are not the same. Data analysis is a subset of data analytics, mainly fasing on past & present dates. On the other hand data analytics is a broader dislipine that includes data analysis along with predicative modeling, automatic and decision making strategies. Data analysis is a fundamental component of business research, necessitating specialized software tools for managing extensive datasets, conducting statistical tests, and executing sophisticated algorithms. Examples include SPSS, R, and Tableau, which we will study and discuss in further detail later in this unit.

13.3 SPSS (STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES)

A lot of people in the social sciences use SPSS, which stands for “Statistical Package for the Social Sciences,” to do statistical research.The first version of SPSS was called the Statistical Package for Social Sciences and was made

by Norman H. Nie, Dale H. Bent, and C. Hadlai Hull. It initially came out in 1968.

SPSS has an easy-to-use interface and is one of the most powerful tools that researchers can use to do complicated statistical tests like ANOVA, factor analysis, and regression. SPSS is especially useful for survey research and social science-based studies because it makes it easier to adjust data, insert data, and present results. Data from almost any type of file can be used by SPSS Statistics to produce tabulated reports, charts, and plots of distributions and trends, as well as descriptive statistics and other complex statistical analyses. Refer figure 13.1 that describe function of SPSS in research.

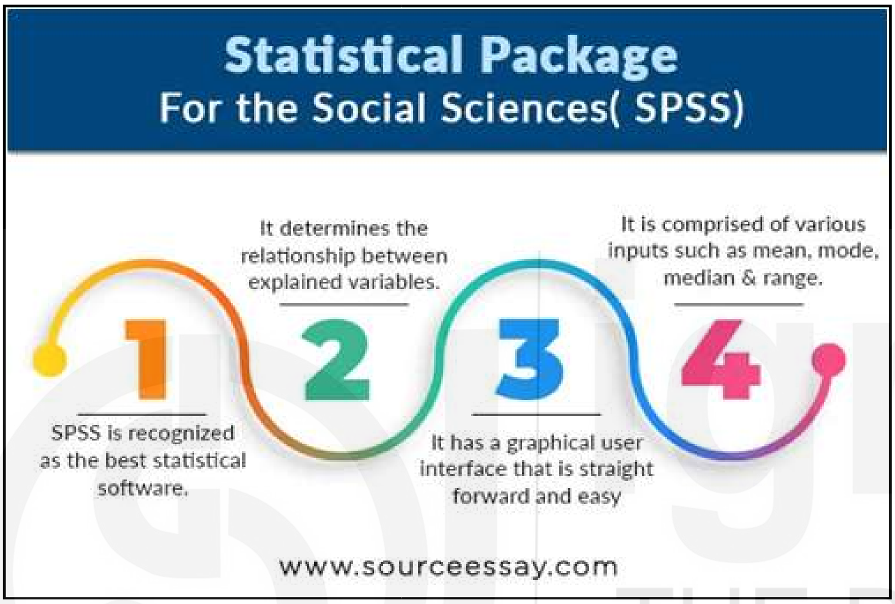


Figure 13.1 : Statistical Package for the Social Sciences (SPSS)

13.4 MICROSOFT OFFICE

Microsoft Office, or simply MS Office, is a suite of productivity software such as client software, server software, and services developed by Microsoft that supports research at various stages. It contains a word processor (Word), a spreadsheet program (Excel) and a presentation program (PowerPoint), an email client (Outlook), a database management system (Access), and a desktop publishing app (Publisher). Several versions of Office cater to various end-users and computing environments.

MS Office makes it easier for people to work more quickly and simply. Word, Excel, and PowerPoint all have features that can help one to get a lot more done, like templates, auto-correct, and advanced editing tools. Such as Word is used to write and edit research papers; Excel is a powerful data analysis tool that allows users to organize, filter, and comprehend data visually with formulas, charts, and pivot tables; and PowerPoint is used to show findings visually.

Figure 13.2: Microsoft Office Applications

MS Office is a versatile software suite that provides essential tools for research and academic writing.

Microsoft Word stands out for its robust features. The Researcher feature in Word enables users to find reliable sources, add content with proper citations, and automatically generate inline citations and bibliography entries. It also offers templates, advanced formatting options, spell and grammar checks, and tools for collaboration and sharing, making it indispensable for drafting and structuring research documents.

Microsoft Excel, another key application in MS Office, plays a critical role in research data analysis. Its powerful functions allow users to organize, calculate, and visualize data efficiently. Researchers can use Excel to create charts, graphs, and pivot tables for a clear representation of findings, making data interpretation more accessible and impactful.

Microsoft PowerPoint is equally essential for presenting research. It enables the creation of visually engaging presentations with features like customizable templates, multimedia integration, and animation tools. PowerPoint helps researchers communicate their findings effectively during conferences, seminars, and academic discussions.

In addition, Microsoft OneNote supports researchers by serving as a digital notebook for organizing ideas, notes, and references in one place. Its ability to integrate with other MS Office tools makes it a seamless part of the research workflow.

In conclusion, MS Office provides a comprehensive suite of applications that cater to various aspects of research and academic writing. From drafting documents and analyzing data to creating impactful presentations and organizing notes, MS Office empowers researchers with the tools they need to succeed.

13.5 R SOFTWARE

R is an open-source programming language and software environment that works well for statistical computing and visualizing data. It can be built and run on a variety of UNIX platforms, including Windows and MacOS. It

employs a variety of statistical techniques, including time series analysis, sorting, grouping, graphical representation, and linear and nonlinear analysis. The fields of data mining, bioinformatics, and data analysis have adopted it.

Professional and academic research extensively uses R due to its flexibility and robust analytical capabilities. Also, it works well for machine learning tools. R is more challenging to learn than Python due to its predominant use in statistics. It might not be as quick as languages like Python, especially when doing tasks that require an excessive amount of computing power or processing a lot of data simultaneously.

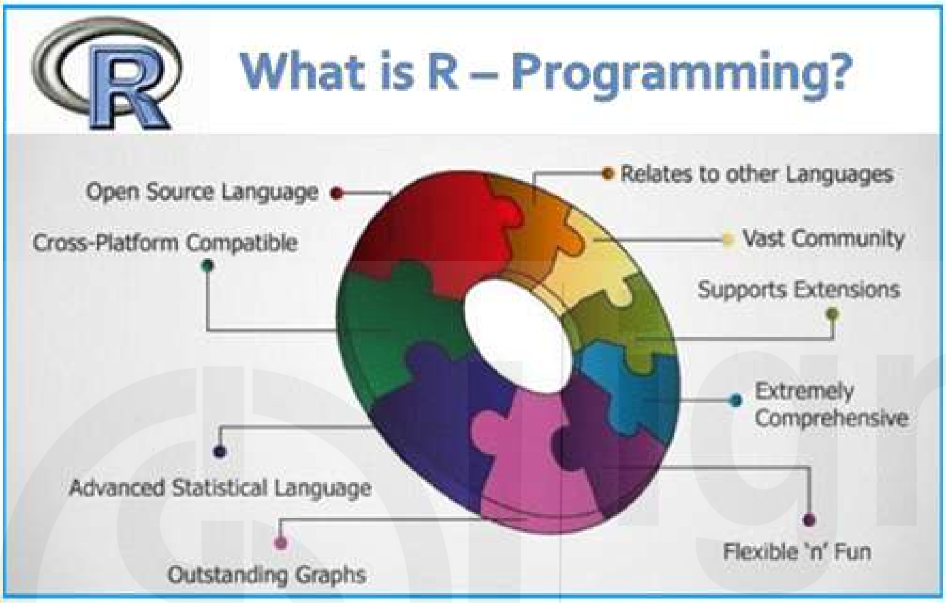


Figure 13.3: R Programming

13.6 TABLEAU

Tableau is visual analytics software that empowers users to create dashboards that facilitate communication and connectivity. People often use Tableau instead of other more standard business intelligence (BI) tools because of its many ways to show data. It works especially well for showing data trends and ideas in a clear and concise way. Its drag-and-drop interface makes complicated data analysis easier and helps experts get their results across clearly.

People can use searches on relational databases, online analysis processing cubes, cloud databases, and spreadsheets to make graph-style data visualizations with Tableau products. To get, keep, and load data again, it can also use an in-memory data engine. Analyst use tableau to work with SQL and extract valuable insights per business leaders. This allows leaders to explore and understand data visually without reading to write code.

Check Your Progress A:

1) **Fill in the blanks:**

   a) Software applications enhance the __________, accuracy, and accessibility of the research process.

   b) Statistical programs such as __________, R, and Python are utilized for complex data analysis in quantitative research.

c) Application software is often referred to as __________ programs because they enable users to complete tasks.

d) Microsoft Office includes a word processor (Word), a spreadsheet program (Excel), and a __________ program (PowerPoint).

2) Name two online survey platforms mentioned that help researchers collect data.

…………………………………………………………………………….

…………………………………………………………………………….

…………………………………………………………………………….

…………………………………………………………………………….

…………………………………………………………………………….

13.7 TOOLS USED IN RESEARCH

We all understand that without good research tools, the process of research would be challenging, time-consuming, and prone to errors. In each and every step of the research process, the research tools are required. Researchers use research tools to help them collect, analyze, and present data. A research tool could be a device, technique, or software. The selection of research tools depends on the research methods, objectives, and needs. The careful selection of research tools help researchers to achieve their goals effectively and efficiently. Some essential research tools include search engines like Google Scholar, JSTOR, and PubMed; reference management software like Zotero, Mendeley, and EndNote; statistical analysis tools like SPSS, R, and Stata; writing tools like Microsoft Word and Grammarly; data visualization tools like Tableau and Excel; and data-gathering research tools like Google Form.

13.7.1 Qualtrics

Qualtrics is an online survey tool that lets users create and send surveys, get answers, and even look at the data from those responses, all from the same platform. Qualtrics boasts numerous features, such as a survey wizard, trial surveys, and integrated graphics and statistical tools. Qualtrics stores the shared data on secure servers, limiting access to only individuals granted permission. Anyone can use Qualtrics on a Mac or a PC, and it works with Chrome, Firefox, Internet Explorer, and Firefox.

13.7.2 Indexed Databases

Indexing is the general term for using a standard sign or measure as an indicator or guideline. In academia, an indexed database is one that uses filters to make it easier for users to find data quickly. Indexed databases are essential for researchers to find academic papers and publications.These databases are required for reviewing articles and gathering data. Some well-known databases are listed below:

13.7.2.1 Google Scholar

Google Scholar offers an efficient method for comprehensive searches of academic literature. From a single location, one may do searches across several disciplines and sources, including articles, theses, books, abstracts, and court verdicts, sourced from academic publishers, professional societies, online repositories, universities, and other websites. Google Scholar assists in locating pertinent research across the global academic landscape. Google Scholar aims to rate papers in accordance with scholars' methodologies, taking into account the complete text of each document, its publication source, authorship, and the frequency and repetition of citations in other academic papers.

Google Scholar has the following features:

- Search all scholarly literature from one convenient place.
- Explore related works, citations, authors, and publications.
- Locate the complete document in your library or online.
- Keep up with recent developments in any area of research.
- Create a public author profile; check who's citing your publications.

13.7.2.2 Scopus

Scopus is a multidisciplinary database that indexes and provides access to peer-reviewed literature, including scientific journals, conference proceedings, books, and trade publications. Scopus covers a wide range of subjects, including science, technology, medicine, social sciences, and arts and humanities. It also includes records from MEDLINE and EMBASE databases.

Scopus is useful for a variety of research tasks, including:

Keyword searching: One can search for articles, conference proceedings, trade publications, and book chapters on a topic.

Citation tracking: To see which publications have cited a particular publication and how many times.

Author metrics: For finding author information, such as their h-index and lists of publications.

Trend analysis: In research, you can analyze trends.

Identifying promising journals: To identify publications that might be good for publishing in.

Finding collaborators: Can search for potential collaborators or subject experts.

13.7.2.3 Web of Science

Web of Science is a citation database that includes academic journals, conference proceedings, and books. It assists academics in identifying novel research domains by analyzing citation trends and evaluating the impact of articles within a specific discipline. This enables researchers to broaden their

interests and discover new subjects. The academic community extensively utilizes Web of Science as an essential instrument for assessing, evaluating, and monitoring scientific research.

13.7.2.4 ProQuest

ProQuest is a multidisciplinary database that is committed to empowering researchers and librarians around the world. Its innovative information content and technologies boost the productivity of students, scholars, professionals, and the libraries that serve them. Through partnerships with content holders, ProQuest preserves rich, vast, and varied information whether historical archives or today's scientific breakthroughs, and packages it with digital technologies that enhance its discovery, sharing, and management. For academic, corporate, government, school, and public libraries, as well as professional researchers, ProQuest provides services that enable strategic acquisition, management, and discovery of information collections. This vast amount of content allows for serious research on virtually any research topic, from multiple perspectives and across multiple formats. ProQuest's renowned ability to structure data for simple access and discovery by students, scholars, and information seekers of all kinds enables precision search of the content.

13.7.2.5 EBSCO

EBSCO is a global leading provider of research databases, e-journal and e-package subscription management, book collection development, and acquisition management, as well as library technology, e-books, and clinical decision solutions for universities, colleges, hospitals, corporations, government entities, schools, and public libraries. EBSCO's finest online full-text databases provide access to articles from peer-reviewed journals published by numerous esteemed publishers worldwide.

13.7.2.6 SCI and Extended

Eugene Garfield originally produced the Science Citation Index (SCI), a citation index. The larger version, Science Citation Index Expanded (SCIE), covers more than 9,200 notable and significant journals across 178 disciplines, from 1900 to the present. SCI-indexed journals are a subset of SCIE-indexed journals and are considered to be more prestigious. They are also more likely to have higher impact factors and cover a wider range of scientific disciplines. A rigorous selection process identifies these as the world's leading journals of science and technology. The small difference between Science Citation Index (SCI) and Science Citation Index ExpandedTM (SCIE) is the storage format. As previously stated, both SCI and SCIE are available online. However, SCI is available on CD/DVD format, while SCIE is not.

13.7.2.7 Gale

Gale is a world leader in e-research and educational publishing for libraries, schools, and businesses. The Gale Literature Resource Centre is a

comprehensive academic database that includes top journals in literature and critical analysis, as well as full-text articles from scholarly journals.

It provides learners with powerful search tools to narrow results and deliver the reliable, timely content they need to be successful researchers. Gale helps libraries connect with researchers looking for credible, timely content. Gale provides thousands of academic resources such as databases, eBooks, primary source content, eLearning solutions, and more.

13.8 CITATION METRICS

Citation metrics play a crucial role in cited reference searching, which involves identifying instances of an article, book, journal, or specific author cited in another work. The number of citations a work receives determines its citation metrics. So, the more citations a work has, potentially, the greater the research's impact. Citation data can be obtained from citation databases, discipline-specific databases, and an emerging range of alternative metrics.

Citation counts are not comparable across disciplines; *for example, in the social sciences and humanities, citation counts are lower because researchers are more likely to publish in books and conference papers that are not well covered by citation databases.* Gathering information about the impact of a work's citations can be beneficial to the researcher in the following ways:

- Citations determine the impact of an article, book, journal, or author in the research field.
- It helps to find appropriate journals to publish in.
- It track the most recent advancements in a specific field.
- It build an impact narrative for performance based research fund (PBRF)
- It identify potential research partners.

13.8.1 Publon

Publons is a platform that provides a free service for academics to track, verify, and showcase their peer review and editorial contributions for academic journals. It helps researchers to build their profile by recording their contributions to the peer review process, which is an important part of academic publishing.

For Example: If you review papers for journals, you can use Publons to keep a record of your reviews. This way, you can show your peers and potential employers your involvement in the academic community and your expertise in evaluating research.

13.8.2 Research Gate

ResearchGate is a European commercial social networking site for scientists and researchers to share papers, ask and answer questions, and find collaborators. ResearchGate is the professional network for scientists and researchers, providing a platform to connect, collaborate, and share your

work with a global community of 25+ million members.

Figure 13.4: Research Gate

13.8.3 Altmetric.com

Altmetrics, or alternative metrics, are a way to measure the impact of research and scholarship online. Altmetrics aim to swiftly measure the impact of research, surpassing traditional citation-based measures like journal prestige, citation counts, and author H-index. Altmetrics aims to provide a comprehensive view of research utilization, taking into account online interactions, social media interactions, and network interactions. Altmetrics can help researchers understand how their work is received by different audiences and devise their own engagement strategies.

Figure: 13.5 Altmetric.com

For Example: Altmetric.com gathers and analyzes information about a research paper's widespread, Twitter mentions, news article features, and blog discussions. It then provides a score, or "Altmetric Attention Score," that reflects the paper's visibility and influence outside traditional academic circles. This aids researchers in comprehending the reception of their work among the general public and other non-academic audiences.

13.9 UNIQUE IDENTIFICATION

In research, unique identifiers serve a variety of purposes by distinguishing one entity from another. Unique identification systems aid in managing and tracking research outputs. It includes the following heads:

13.9.1 ORCID (Open Researcher and Contributor ID)

ORCID (Open researcher and contributor ID) is a unique identifier for researchers. ORCID is a free, unique, persistent identifier (PID) for individuals to use as they engage in research, scholarship, and innovation activities. It aids in differentiating researchers from each other and guarantees

accurate attribution of their work. Think of it as a personal ID number for researchers, like a Social Security number for researchers' academic work.

For Example: Imagine you are Mr. X, a researcher. There are many researchers with the same name, so it might be hard to know if a paper you're reading was written by your Mr. X or someone else with the same name. If Mr. X has an ORCID ID, it will look something like this: 0000-0002-1825-0097. When Mr. X publishes his papers, they include his ORCID ID, ensuring that everyone knows exactly who wrote them. This helps to accurately track their research work and contributions.

13.9.2 DOI

A DOI (Digital Object Identifier) is a unique link or code that helps you find and access a specific digital document, like a research paper, online. It's like a permanent web address for that document, ensuring you can always locate it even if the web page changes.

For example: You can directly access a research paper, regardless of its hosting location, by entering its DOI, such as 10.1234/example, into a search tool or website.

13.9.3 ISSN and eISSN

An eight-digit serial number, known as an International Standard Serial Number (ISSN), uniquely identifies a serial publication (periodical), like a magazine. The e-ISSN, also known as the Electronic ISSN, serves as the International Standard Serial Number (ISSN) for online versions of serial publications. The ISSN is especially helpful in distinguishing between serials with the same title.

eISSN 97729484332 9 772948 433202	**eISSN (Electronic International Standard Serial Number)** is the ISSN specifically for electronic versions of serial publications, like online journals or digital magazines.
ISSN 97728054520 9 772805 452001	**ISSN (International Standard Serial Number)** is a unique 8-digit code used to identify serial publications like journals and magazines. It helps to standardize and differentiate between various serial publications.

Figure 13.6 ISSN Number

13.10 REFERENCE MANAGEMENT TOOLS

Reference management tools are software applications designed to help researchers and writers to organize and manage their references and bibliographic information. They make it easier for researchers to collect, store, and format references for academic papers, theses, and other research projects.

Some of the key features of reference management tools are as follows:

1. **Collecting References:** The tool allows you to directly save references from databases, journals, and websites.
2. **Organizing References:** Sort and categorize references into folders or groups.
3. **Generating Citations and Bibliographies:** Automatically format citations and bibliographies in various styles (like APA, MLA, or Chicago).
4. **Sharing and Collaboration:** Share reference libraries with collaborators and work together on research projects.

Some reference management softwares are:

13.10.1 Mendeley

In 2007, PhD students Paul Foeckler, Victor Henning, and Jan Reichelt established Mendeley, which was subsequently acquired by the Dutch academic publishing company Elsevier in 2013. Researchers use it to manage and share research papers, as well as to generate in-text citations and bibliographies for scholarly articles.

MENDELEY

Figure 13.7 Mendeley

Mendeley is based on a web service that requires users to first create a personal account on the Mendeley site. On the local computer side, it works in conjunction with a word processor plug-in or add-in (mandatory) and a desktop application (now facultative for the latest current version). The new "Mendeley Reference Manager library" is cloud-based. This ensures that the cloud always stores the most recent changes and makes references accessible across multiple devices.

13.10.2 Zotero

Zotero is an open-access, user-friendly reference management tool that functions as a personal research assistant for researchers or academicians, assisting in the collection, organization, citation, and sharing of your research

sources. Zotero runs on several operating systems, including Windows, Mac and Linux.

Figure 13.8: Zotero

Zotero allows researchers and academicians to:

1. Save references from library catalogs, research databases, and the Web.
2. Add PDFs, images, audio and video files, snapshots of web pages, and more.
3. Write annotations and attach them to citations.
4. Create bibliographies using major citation styles.

13.11 ETHICAL CONSIDERATIONS IN RESEARCH

Most people learn ethical norms at home, at school, in church, or in other social settings. Although most people acquire their sense of right and wrong during childhood, moral development occurs throughout life, and human beings pass through different stages of growth as they mature. But in the context of research, ethical considerations are a set of principles that guide research practices and designs. Software usage in research must comply with ethical norms. Some ethical considerations include these principles: voluntary participation, informed consent, anonymity, confidentiality, potential for harm, results communication, plagiarism, data privacy, and intellectual property rights are paramount. The ethical application of databases, citation management tools, and data analytics software is essential for maintaining research integrity and academic honesty. We will discuss this in detail in next Unit.

Check Your Progress B:

1) Differentiate between research tools and statistical analysis tools.

   …………………………………………………………………………………

   …………………………………………………………………………………

   …………………………………………………………………………………

   …………………………………………………………………………………

2) State true or false for the following statements:

   a) Qualtrics is primarily used for statistical analysis.

   b) Citation counts are comparable across all disciplines.

   c) ORCID is a unique identifier specifically for academic journals.

   d) Google Scholar allows users to search for academic literature across multiple disciplines from a single location.

13.12 LET US SUM UP

In today's research landscape, software plays a crucial role in making the process more efficient, accurate, and accessible. This unit explores different types of software used in business research, from tools that help collect data to those that analyse and present findings.

We begin by categorizing software into three types: system software, application software, and data analytical software. System software, like Windows and macOS, supports the operation of research tools. Application software, such as Microsoft Office, web browsers, and project management tools, helps with writing reports, organizing data, and managing research project. Data analytical software, like SPSS, R, and Tableau, assists in processing and analysing large datasets, making it cashier to find patterns and insights.

The unit also highlights tools researches rely on, such as survey platforms (Qualtrics, Google Forms) for data collection and indexed databases (Google Scholar, Scopus, Web of Science, ProQuest, and EBSCO) for finding academic papers. Citation metrics like Publons, Research Gate, and Altmetric.com help researchers measure the impact of their work. Unique identifiers like ORCID, DOI, and ISSN ensure proper recognition of research contributions.

To keep references organized, tools like Mendeley and Zotero simplify citation management and make collaboration casier. The unit also emphasizes the ethical aspects of research, including plagiarism prevention, data privacy, and intellectual property rights.

In short, research software is not just about numbers and statistics, it's about making the research process smoother, improving collaboration, and ensuring that findings are presented clearly and ethically. By using the right tools, researchers can save time, avoid errors, and focus on what really matters is generating meaningful insights.

13.13 KEYWORDS

SPSS : is useful for survey research and social science-based studies because it makes it easier to adjust data, insert data, and present results.

Indexed Databases: are essential for researchers to find academic papers and publications.These databases are required for reviewing articles and gathering data.

ORCID : is a unique identifier for researchers. ORCID is a free, unique, persistent identifier (PID) for individuals to use as they engage in research, scholarship, and innovation activities.

DOI : is a unique link or code that helps you find and access a specific digital document, like a research paper, online.

13.14 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress A:

1) a) efficiency b) SPSS c) productivity d) presentation

Check Your Progress B:

1) a) False b) False c) False d) True

13.15 TERMINAL QUESTIONS

1) What role do software applications play in enhancing the research process?
2) What is the purpose of reference management tools in the research process?
3) Describe the main features of SPSS that make it suitable for statistical research.
4) How does Qualtrics enhance the survey creation process for researchers?
5) What is the primary function of indexed databases in academic research?
6) Explain the difference between DOI and ORCID in the context of research identification.

13.16 FURTHER READINGS

1. Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., & Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International journal of information management*, *57*, 101994.
2. Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: how may AI and GPT impact academia and libraries?. *Library hi tech news*, *40*(3), 26-29.
3. Fry, L. W. (1982). Technology-structure research: Three critical issues. *Academy of Management Journal*, *25*(3), 532-552.

Note	These questions are helpful to understand this unit. Do efforts for writing the answer of these questions but do not send your answer to university. It is only for your practice.

UNIT 14 ETHICAL ISSUES IN RESEARCH

Structure

14.0 OBJECTIVES

After studying this unit, you will be able to:

- Explain the concept of ethics in research;
- Emphasize the importance of informed consent in ensuring participants' autonomy;
- Identify and provide solutions to common ethical issues in research;
- Explain ethical guidelines for different stakeholders in business research: and
- Define plagiarism and its consequences.

14.1 INTRODUCTION

In the previous unit, we explored the transformative role of technology and software in research, emphasizing their pivotal contributions to enhancing efficiency, accuracy, and scalability. Tools such as data analysis platforms, statistical software, and automated data collection technologies have revolutionized how researchers gather, process, and interpret data. From facilitating real-time data collection through digital surveys to leveraging

software for complex statistical modeling, the integration of technology has become indispensable in modern research practices. However, with these advancements come significant responsibilities. As researchers adopt sophisticated tools for data collection and analysis, ethical considerations become paramount. The ease of accessing vast amounts of information, often including sensitive or personal data, raises questions about privacy, consent, data integrity, and the potential for misuse. Ethical guidelines serve as the compass to navigate these challenges, ensuring that technological progress aligns with principles of fairness, respect, and accountability.

Since research involves investigation, collection, interpretation and documentation, it becomes important that the researchers adhere to the defined protocol. Past studies have emphasized that conducting business research requires a professional and responsible approach. Data collection should be carried out with the consent of respondents, ensuring ethical and controlled practices. Interpretation must be handled with utmost care. Many organizations have developed their own codes of ethics to guide research practices. While defining business ethics, including research ethics, is common in Western organizations, in India, such guidelines are primarily documented in sectors such as pharmaceuticals and certain banks like HSBC. Additionally, international bodies provide comprehensive ethical standards, including the Social Research Association's (SRA) ethical guidelines, the American Psychological Association (APA) code of ethics, the Council of American Survey Research Organizations' (CASRO) code of standards and ethics for survey research, and the codes of conduct and ethics established by the American Marketing Association (AMA) and the Business Marketing Association (BMA).

In ethics, or moral philosophy, conceptions of right and wrong behavior are defended, systematized, and minimized. It is a branch of philosophy that has standards, rules, or value systems. Ethical issues in research are some of the rules that researchers follow to ensure protecting the rights in formulating research strategies and establishing a trusted relationship between the investigator and study participants. Research ethics refers to a persistent experience towards producing new information and knowledge for the purpose of communicating a new stream of ideas in academia and plays an important role in a scientific study.

14.2 IMPORTANCE OF ETHICS IN RESEARCH

Since researcher work in a dynamic global research ecosystem with multiple stakeholders with diverse interests, it is necessary for them to maintain research integrity, based on a set of shared values that include ethics, relevance, transparency, respect and accountability etc. Hence, the role of morals and ethics becomes important. The word Ethics implies a system of accepted beliefs that control our behavior, especially when a system that is based on morals. The word Morality on the other hand, implies a set of personal or social standards for good or bad behavior and character or the quality of being right, honest or acceptable. The importance of ethics can be understood from below point:

1. **Protecting Participant Safety:** Ensuring the safety and well-being of research participants is a fundamental aspect of research ethics. Researchers must prioritize minimizing physical, psychological, or emotional harm to participants. This includes obtaining informed consent, maintaining confidentiality, and avoiding coercion or undue influence.

2. **Maintaining Scientific Integrity:** Ethical research practices uphold the credibility and reliability of scientific findings. By ensuring accurate data collection, analysis, and reporting, researchers maintain the integrity of their work and prevent issues like falsification, fabrication, or plagiarism. This is critical for advancing knowledge and fostering trust in research.

3. **Upholding Human Rights and Dignity:** Respecting the rights and dignity of all participants is central to ethical research. Researchers must treat participants with fairness and respect, ensuring they have the autonomy to make informed decisions about their involvement. Special care is needed when working with vulnerable populations to avoid exploitation.

4. **Ensuring Social Responsibility:** Ethical research contributes to societal well-being by addressing relevant issues and providing solutions to global challenges. Researchers have a responsibility to avoid causing harm to society, whether through biased findings, misuse of results, or unethical practices. Socially responsible research aligns with the greater good.

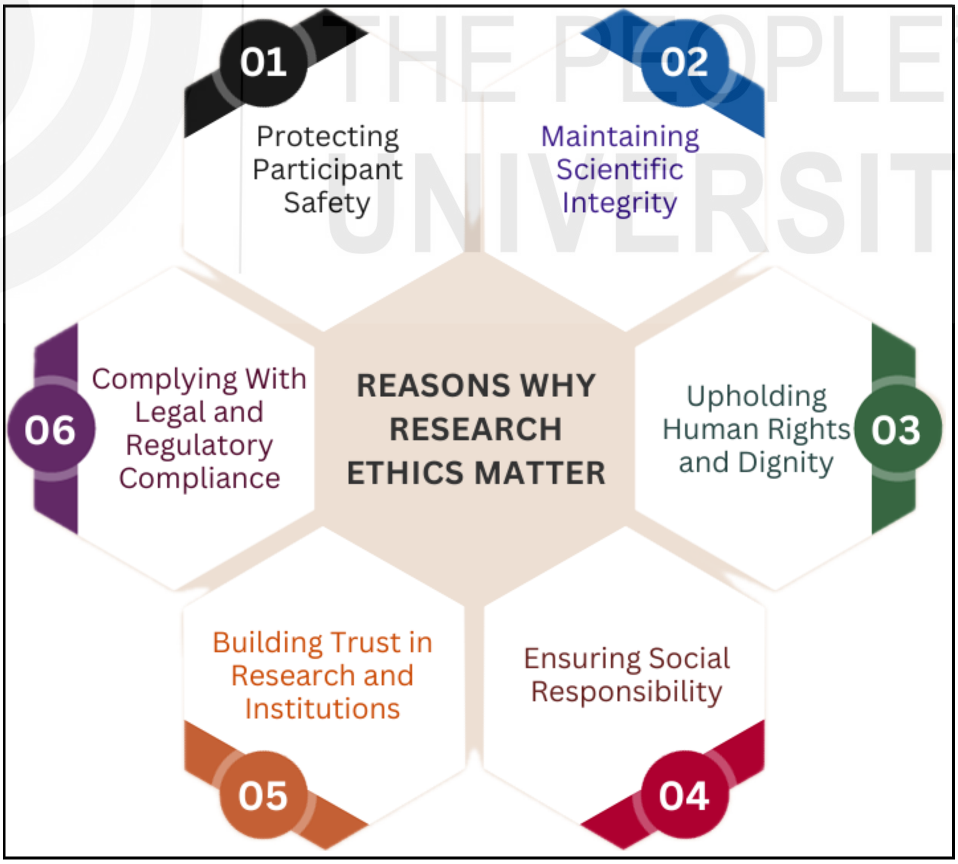


Figure 14.1: Importance of Ethics in Research

5. **Building Trust in Research and Institutions:** Trust is essential for the success and acceptance of research. When researchers adhere to ethical principles, they build confidence among participants, institutions, and the broader community. This trust encourages participation, supports funding opportunities, and enhances the reputation of research organizations.

6. **Complying with Legal and Regulatory Compliance:** Research ethics often overlap with legal and regulatory frameworks designed to protect participants and ensure fair practices. Adherence to these laws and guidelines, such as data protection regulations (e.g., GDPR) and institutional review board (IRB) standards, is crucial for conducting legitimate and compliant research.

14.3 ETHICAL ISSUES IN RESEARCH

Ethical issues in research play a critical role in maintaining the credibility, reliability, and societal value of scientific endeavors. Researchers must navigate a range of ethical challenges to ensure that their work upholds the principles of honesty, fairness, and respect for all stakeholders involved. Below, we explore the key ethical issues in research, as highlighted in the Figure 14.2:

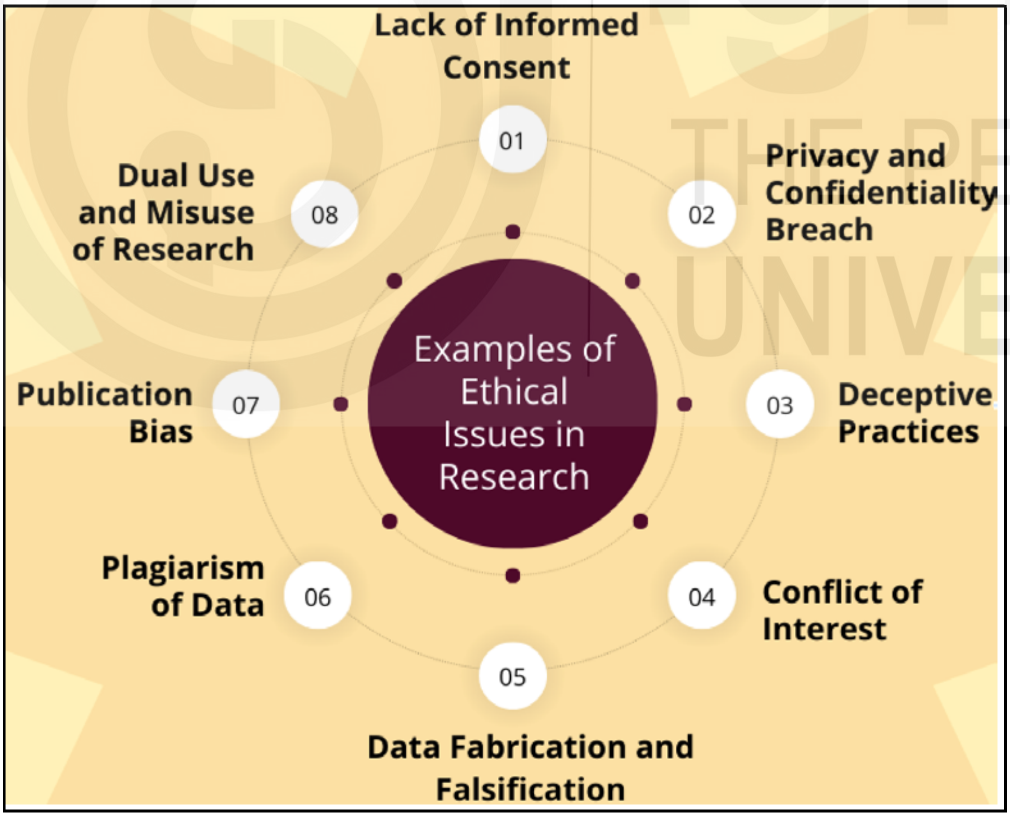


Figure 14.2 Common Ethical Issues in Research

1. **Lack of Informed Consent:** One of the most fundamental ethical principles in research is obtaining informed consent from participants. Failing to do so violates their autonomy and can result in harm or discomfort. Informed consent ensures that participants are fully aware of the nature, purpose, risks, and benefits of the study, and that they

voluntarily agree to participate. We will discuss in detail about this in below section 14.4.

2. **Privacy and Confidentiality Breach (Anonymity and confidentiality):** Protecting the privacy of participants and ensuring the confidentiality of their data are critical ethical obligations. A breach of privacy can lead to emotional distress, reputational damage, or even legal consequences for participants. Researchers must adopt secure data management practices and seek explicit consent for data sharing or publication. This will be discussed in details in section 14.5.

3. **Deceptive Practices:** Deliberately misleading participants or withholding critical information undermines the ethical foundation of research. Deception, unless absolutely necessary and justified (e.g., in certain experimental designs), erodes trust and compromises the integrity of the research process. This will be discussed in detail in below section 14.6.

4. **Conflict of Interest:** Conflicts of interest arise when a researcher's personal, financial, or professional interests interfere with their ability to conduct unbiased research. Such conflicts can compromise the validity of findings and damage the credibility of the research. Full disclosure of potential conflicts is essential to mitigate their impact.

5. **Data Fabrication and Falsification:** Fabricating or falsifying data is a severe ethical violation that undermines the foundation of scientific knowledge. Data fabrication involves inventing data, while falsification refers to altering data to fit desired outcomes. Both practices mislead the scientific community and harm public trust in research.

6. **Plagiarism of Data:** Plagiarism involves using someone else's work, ideas, or data without proper acknowledgment. This not only disrespects the original creator but also violates the ethical standards of academic and scientific integrity. Researchers must always provide appropriate citations and references for borrowed content.

7. **Publication Bias:** Publication bias occurs when researchers selectively report positive or significant findings while ignoring negative or inconclusive results. This creates a distorted picture of the research landscape and hinders the progress of evidence-based knowledge.

8. **Dual Use and Misuse of Research:** Research intended for beneficial purposes can sometimes be misused for harmful or unethical objectives, such as the development of weapons or harmful technologies. Researchers must carefully evaluate the potential consequences of their work and ensure it aligns with societal welfare.

14.4 INFORMED CONSENT

Informed consent is a cornerstone of ethical research, ensuring that participants voluntarily agree to participate in studies with full awareness of their purpose, risks, and benefits.

In the context of business research, informed consent plays a critical role in maintaining ethical standards. *For instance, in employee satisfaction surveys, employees must be informed about the study's objectives, assured of confidentiality, and allowed to participate voluntarily without fear of repercussions*. Similarly, *in customer behavior studies, collecting personal data without consent can violate ethical norms and legal requirements, such as those under GDPR, potentially damaging the organization's reputation*. Transparency about how collected data will be used is crucial to building trust and ensuring compliance with regulations.

To uphold ethical business research, researchers must clearly communicate the study's purpose, such as improving workplace policies or enhancing customer experience, using language that participants can easily understand. They should respect participants' privacy by obtaining consent in a non-coercive manner and addressing any questions or concerns openly. For sensitive studies, such as those on employee mental health or consumer financial behavior, researchers must handle topics carefully to avoid causing emotional discomfort or distress. Even with consent, participants should not feel pressured to disclose sensitive information, and safeguards must be in place to protect their confidentiality and dignity.

In cases involving sensitive data, such as biometric screenings or health assessments, written formal consent should be obtained to ensure participants are fully aware of the implications of their involvement. Another example in today's contemporary context is "***A leading e-commerce platform implemented AI-driven personalized marketing based on customer purchase history. The company informed customers about the use of AI to analyze their data and offered an option to opt out of the program. By providing clear terms of participation and guaranteeing data privacy, the platform not only complied with legal standards but also enhanced customer trust and engagement***".

Ethical considerations must be at the forefront of all research activities, ensuring the integrity of the study and fostering trust among participants. By adhering to these principles, business research can align with both ethical and legal standards, ultimately benefiting participants and advancing meaningful outcomes.

14.5 ANONYMITY AND CONFIDENTIALITY

In business research, maintaining anonymity and confidentiality is crucial to ensure ethical integrity and protect the rights of participants. Business research often involve sensitive topics such as employee satisfaction, workplace harassment, consumer financial behavior, or whistleblower reports, where revealing participant identities can lead to significant personal or professional consequences. Researchers must take stringent measures to ensure that participants' identities and responses remain protected.

For example, in an employee engagement survey, employees may provide feedback on sensitive issues like toxic workplace culture or unfair practices. To encourage honest responses, the survey must guarantee anonymity. This

can be achieved by assigning each participant a code number instead of using their real names and ensuring that no identifiable information is included in the analysis or reports. Similarly, in customer satisfaction research, participants might share personal details about their purchasing behaviors or financial constraints. These details must be anonymized, ensuring that even researchers cannot link responses back to specific individuals.

Anonymity is described as a situation where "neither the researchers nor the readers of the findings can identify a given response with a given respondent." To implement this, researchers should avoid recording names or any identifying details in survey forms or interview notes. Instead, respondents can be assigned pseudonyms or unique codes, and only aggregate data should be presented in reports. If interviews are recorded, the researcher must take extra precautions to protect the recordings and ensure that identifying information is never disclosed, even inadvertently, during reporting or publication.

While anonymity prevents the linking of responses to individuals, confidentiality ensures that researchers, though aware of the identity of respondents, do not disclose this information publicly. *For instance, in a market research study on consumer behavior for a luxury product, some participants may not want their financial status or spending habits revealed. Though researchers may need to know participant details to validate data, they must promise and ensure that this information will not be disclosed to any third party.*

A practical example of the challenges of confidentiality in business research can be seen in whistleblower studies. Suppose a company is conducting an internal investigation into workplace harassment. If employees fear their identities will be exposed, they may refrain from participating. The researcher must explicitly guarantee confidentiality and ensure that responses are not shared with management in a way that could identify the whistleblower.

To ensure anonymity and confidentiality, business researchers should adhere to the following best practices:

- Avoid including names or identifying details in survey forms or interview notes.
- Use code numbers or pseudonyms for participants.
- Store questionnaires, recordings, and notes securely to prevent unauthorized access.
- Do not discuss participant information or responses with colleagues, family members, or third parties.
- Ensure that reports or publications exclude any details that could inadvertently reveal participant identities.
- Seek participant consent in writing, clarifying how their information will be used and protected.

An illustrative example of these principles in action is a customer research project on loan repayment difficulties. Respondents may share intimate details about their financial struggles, which, if exposed, could harm their personal and professional lives. Even when collecting this data anonymously, the researcher must be cautious about how the data is stored and reported. Summarizing findings at a group level, without including direct quotes or identifiable anecdotes, helps maintain both anonymity and confidentiality.

In some cases, protecting confidentiality may even require ethical sacrifices, such as withholding data from authorities. *For example, during a corporate whistleblower investigation, a researcher might face pressure to disclose participant details*. Upholding confidentiality, even in such challenging circumstances, is essential to maintaining trust and ensuring ethical research practices. By prioritizing anonymity and confidentiality, business researchers can encourage honest participation, protect respondents from harm, and ensure the ethical credibility of their studies.

14.6 DECEPTION

Deception in research refers to the practice of concealing the true identity or objective of a study from participants. This ethical dilemma arises when researchers justify such practices by focusing on the potential benefits of the data collected. For instance, *A student conducting research on reproductive health faced significant challenges when respondents refused to engage with her, stating that, as an unmarried woman, she would not understand their problems. To overcome this, she introduced herself as a married mother of two children*. This deceit allowed her to gather detailed and high-quality data, as respondents became comfortable seeking her opinion on sensitive issues such as health problems, child spacing, and birth control. While the researcher believed the deception was harmless and beneficial, it raises critical ethical questions about the morality of concealing one's true identity.

Such an approach aligns with the theoretical framework of consequentialism, which argues that actions are justified if they result in greater benefits. In many experimental research, participants are often kept unaware of the true purpose of the research, as researchers believe this prevents bias and ensures authentic responses.

Globally, researchers often contend that deception is sometimes unavoidable to achieve research objectives. In such cases, debriefing—explaining the true purpose of the research after its completion—is recommended as an ethical remedy. Debriefing involves revisiting participants to clarify the objectives of the study and assess any potential harm caused by the deception.

Deception in business research involves concealing the true objective of a study or misrepresenting aspects of the research to participants. While such practices can sometimes produce valuable insights, they raise critical ethical concerns. *For instance, a researcher conducting a study on workplace culture might misrepresent the purpose of the study to avoid biased responses from employees. Telling participants that the research focuses on improving*

organizational policies, rather than explicitly stating that it is about workplace harassment, might encourage more honest feedback. While the intention is to gather authentic data, it constitutes deception as the true purpose of the study is not disclosed.

While proponents of deception argue that debriefing neutralizes harm, it is not without its drawbacks. For some participants, learning about the deception after the study may create feelings of doubt, mistrust, or psychological discomfort. Respondents may question their own behavior or worry about how their responses were interpreted. Pat research aptly notes that while "debriefing can be effective in easing ethics in research, the discomfort caused during a study or experiment involving deception is insufficient to fully reverse negative feelings experienced by those research subjects who are prone to having negative feelings about themselves."

Therefore, although deception may occasionally produce valuable insights, it is crucial to weigh its merits and demerits carefully. Researchers have a responsibility toward their participants and must prioritize transparency wherever possible. When deception is unavoidable, it should be implemented with great care, and thorough debriefing should follow to mitigate potential harm. Ultimately, researchers must ensure that the trust and dignity of participants remain central to their work.

Check Your Progress A:

1) Explain some unethical practices in research.

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

2) What is Deception?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

3) What is the importance of Ethics in Research?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

14.7 REPORTING AND FEEDBACK

Over the years, one of the most overlooked aspects of research has been the responsibility of researchers to report their findings back to the communities or populations they studied. Traditionally, researchers would use their

findings to publish articles in peer-reviewed journals or submit reports to institutions, often neglecting to share these results with the people who were directly involved. While some findings might make their way to local or national media outlets, they were rarely communicated effectively to the communities that provided the data. This lack of feedback creates an ethical gap, as these populations remain unaware of findings that could significantly impact their lives.

For example, consider a researcher investigating iron or iodine deficiencies within a population. If the findings are not shared with the affected community, they continue to suffer from these deficiencies without any awareness or guidance on how to address the problem. Such negligence is now widely regarded as unethical under contemporary research guidelines, which emphasize the moral obligation of researchers to disseminate their findings to participants and the broader community.

In social science research, this process can sometimes lead to challenges, especially when the findings are controversial. For instance, early studies on drug abuse in Punjab faced backlash from the community, as they believed the findings tarnished their reputation by unfairly targeting the Punjabi population. In such situations, researchers must remain calm and empathetic, explaining the methods and rationale behind their findings. Additionally, it is advisable to engage community leaders and disseminate findings through local gatherings, such as village meetings, panchayats, or community forums. This approach not only fosters trust but also ensures that findings are understood in the right context.

A best practice in ethical reporting involves first sharing findings with the concerned community, gathering their feedback, addressing any criticisms, and then presenting the results to the broader academic or professional community. Allowing participants and their representatives to review the data ensures transparency and builds mutual trust. It also provides researchers an opportunity to refine their methodology or conclusions based on community input before disseminating findings to a wider audience.

In conclusion, reporting and feedback are integral components of ethical research. By prioritizing communication with the populations involved and incorporating their feedback, researchers can ensure that their work not only advances knowledge but also serves the communities they study in meaningful and respectful ways.

Check Your Progress B:

1) Explain the importance of informed consent in research and provide examples of its application in sensitive studies.

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

2) Discuss the ethical challenges associated with anonymity and confidentiality in business research.

……………………………………………………………………………….

……………………………………………………………………………….

……………………………………………………………………………….

……………………………………………………………………………….

3) What is deception in research, and under what conditions can it be justified? Provide examples.

……………………………………………………………………………….

……………………………………………………………………………….

……………………………………………………………………………….

……………………………………………………………………………….

4) Choose the correct option

i) What does informed consent ensure in research?

a) Anonymity of participants

b) Voluntary and informed participation

c) Guaranteed positive outcomes for participants

d) The absence of ethical guidelines

ii) Which ethical principle focuses on fairness in participant selection and benefit distribution?

a) Confidentiality

b) Anonymity

c) Justice

d) Beneficence

iii) What is the primary purpose of debriefing in research?

a) To collect additional data from participants

b) To explain the true purpose of the study after completion

c) To obtain informed consent

d) To ensure anonymity

iv) What is one of the main challenges in maintaining confidentiality in research?

a) Avoiding informed consent

b) Protecting participant data from unauthorized access

c) Ensuring a high response rate

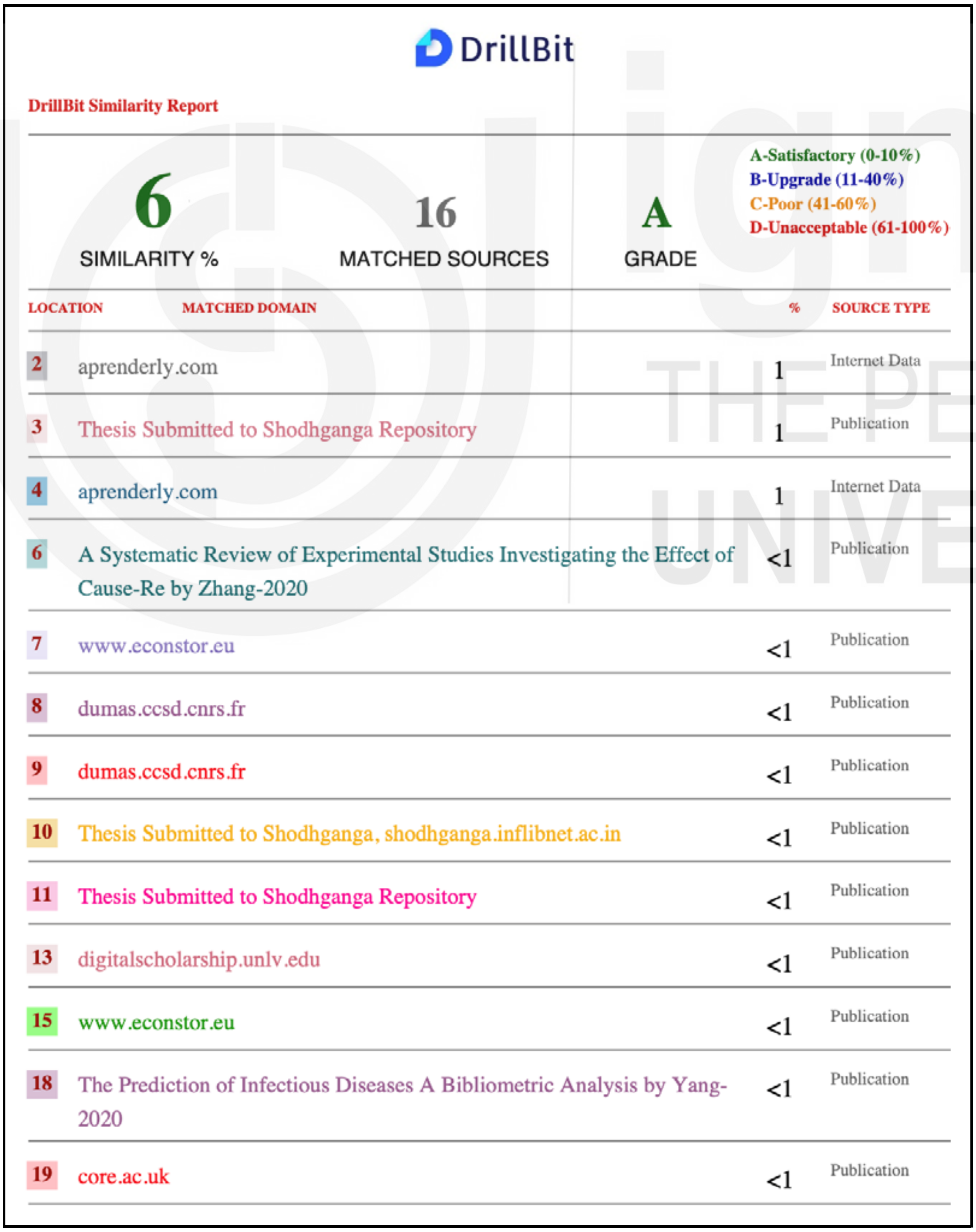
d) Providing financial compensation to participants

14.8 PLAGIARISM AND ITS IMPACT

Plagiarism means using someone else's work, ideas, or intellectual property without proper permission or attribution and presenting it as his/her own work. For instance, Mr. A incorporated a large portion of Mr. B's text and data into his dissertation without proper acknowledgement and presented it as

his own findings. The most common type of plagiarism is using someone else's work, either published or unpublished, and claiming authorship. Generally, the plagiarized material is not from one source only but taken from various sources. Citing the passages from other's texts without proper references also comes under plagiarism. There are other types of plagiarism, including using unpublished data and text from someone else's work for analysis and writing. For instance, student interns are using the records of the institution without permission. All such types of plagiarism must be prevented.

Plagiarism can have severe personal, legal, professional, and ethical impacts (refer Table 14.1). Plagiarists are being caught with the use of plagiarism detection software so readily available like Turntin and Drillbit (refer Figure 14.3 for plagiarism report given by drill bit software). Plagiarists may include academics, students, professionals, authors, journalists, and others. A person, once accused of plagiarism, will probably always be viewed with suspicion.



DrillBit

DrillBit Similarity Report

6	16	A	A-Satisfactory (0-10%) B-Upgrade (11-40%) C-Poor (41-60%) D-Unacceptable (61-100%)
SIMILARITY %	MATCHED SOURCES	GRADE	

LOCATION	MATCHED DOMAIN	%	SOURCE TYPE
2	aprenderly.com	1	Internet Data
3	Thesis Submitted to Shodhganga Repository	1	Publication
4	aprenderly.com	1	Internet Data
6	A Systematic Review of Experimental Studies Investigating the Effect of Cause-Re by Zhang-2020	<1	Publication
7	www.econstor.eu	<1	Publication
8	dumas.ccsd.cnrs.fr	<1	Publication
9	dumas.ccsd.cnrs.fr	<1	Publication
10	Thesis Submitted to Shodhganga, shodhganga.inflibnet.ac.in	<1	Publication
11	Thesis Submitted to Shodhganga Repository	<1	Publication
13	digitalscholarship.unlv.edu	<1	Publication
15	www.econstor.eu	<1	Publication
18	The Prediction of Infectious Diseases A Bibliometric Analysis by Yang-2020	<1	Publication
19	core.ac.uk	<1	Publication

Figure 14.3 Plagiarism report

Here are some of the consequences of plagiarism:

1. **Destroyed Student Reputation**: A plagiarism accusation may cause a student to be expelled or suspended. Their academic record disclosing ethical offenses may cause the student to be barred from joining college from high school or another college. Academic institutions such as schools, colleges, and universities consider plagiarism seriously.

2. **Destroyed Professional Reputation**: A businessperson, public figure, or politician may recognize that the loss from plagiarism follows them for their entire professional lives. They will not only likely be asked to step down or fired from their present position on the job, but they will certainly find it difficult to get another respectable job.

3. **Destroyed academic reputation**: The impacts of plagiarism have been widely reported in academia. An academic's career can be ruined, once scarred with plagiarism allegations. Publishing the manuscript is an essential component of an academic career. Plagiarism may cause losing the capacity to publish, which indicates the end of one's academic career.

4. **Legal Repercussions:** The legal repercussions of plagiarism can be serious. One cannot use another's work without acknowledgement. Copyright laws are absolute and an author has the legal right to sue a plagiarist. There are some forms of plagiarism that may also be deemed a criminal offense, perhaps leading to a prison sentence. Those whose livelihood solely depends on writing, such as journalists or authors, are especially susceptible to plagiarism issues. Writers are very well known about copyright laws and various ways to avoid plagiarism. As a professional writer, plagiarizing is a serious ethical and legal issue.

5. **Monetary Repercussions**: In recent times, many news reports and articles have disclosed plagiarism by journalists, researchers, authors, and public figures. In the case where a journalist works for a newspaper, magazine, or other publisher, or even if a school student is found guilty of plagiarizing, the plagiarist may have to pay monetary penalties.

6. **Loss of Professional Licenses or Certifications**: Plagiarism can lead to disciplinary action from licensing boards or regulatory authorities for professionals working in regulated sectors. For instance, in academia if professors and researchers found plagiarizing could lose tenure or face bans from academic publishing, If lawyers found guilty of plagiarizing, they may face the risk of suspension or disbarment, and if medical professionals are found plagiarizing patient care protocols or research papers, they may face suspension of their medical licenses. A person's career is put on hold and their reputation is harmed when they lose their professional license, making it almost hard for them to return to the industry.

7. **Harm to Organizational Reputation**: Plagiarism by an employee may damage the reputation of the organization. For instance, a researcher who has been plagiarized in a well-known work, their affiliated university or organization may suffer as a result, a marketing firm who is suspected of

copying a campaign from another company might lose customers and face public criticism, and a journalist found guilty of plagiarizing may damage the credibility of their publication and make readers question the reliability of its reporting.

8. **Emotional and Professional Isolation**: Plagiarising may lead to professional and social isolation as colleagues and peers may view the plagiarist as untrustworthy or unethical, so they distance themselves from such individuals. This separation can make it difficult to network, collaborate, or rebuild a career. Being publicly identified as a plagiarist can have a negative emotional impact on motivation and confidence, which can further impede professional recovery.

9. **Retractions and Public Apologies**: In fields like academia and publishing, plagiarised work is subject to official retraction. Retractions may include public statements admitting misconduct may have long-term effects on a person's reputation. Also, plagiarists may be required to issue apologies to the original creator and affected parties either privately or publicly. These actions not only have legal binding but also serve to alleviate the damage to reputation caused by plagiarism.

Table 14.1 Consequences of Plagiarism in Thesis

Similarity index	Consequences of Plagiarism
0-10%	• No penalties
10%-30%	• Revise and resubmit, bring down the similarity index below 10%
30%-50%	• May be asked to withdraw manuscript. • Barred from submitting the script for the year • The Penalty of degrading the grade
>50%	• Barred from submitting the scripts for three years • A fail grade may be awarded

14.9 ETHICAL GUIDELINES IN BUSINESS RESEARCH

Ethics in business research forms the foundation for credible, responsible, and transparent study practices. The principles laid out by associations like the American Marketing Association (AMA) and the Business Marketing Association (BMA) emphasize adherence to ethical standards to protect the interests of all stakeholders involved. These stakeholders include:

1. The sponsoring clients or decision-makers.

2. The respondents from whom information is gathered.

Each stakeholder group has distinct interests and needs, leading to unique ethical considerations. Addressing these concerns effectively ensures the

integrity and success of the research process. These considerations align closely with common ethical challenges in research, including issues of anonymity, confidentiality, deception, and plagiarism, which were discussed earlier. The sections below outline ethical guidelines specified to each group, emphasizing solutions to these challenges with practical examples and best practices.

14.9.1 Ethical Considerations for Decision Makers

Research, like any business transaction, involves an exchange between various entities, primarily the sponsoring client or decision maker and the researcher. Both parties carry ethical responsibilities to ensure the integrity and reliability of the research process. **Following are the ethical concern which client or decision maker must take into consider:**

1. **Objectivity and Transparency:** Decision maker must prioritize objectivity in acquiring and interpreting information. However, conflicts can arise when personal or organizational interests interfere. *For example, a marketing manager who has connections with a particular advertising agency might push researchers to recommend that agency, even if the data suggests better options, to benefit their personal interests.* Such actions compromise the study's validity and ethical standards.

2. **Unethical Use of Research Proposals:** It is common for small or emerging firms to request detailed proposals from research agencies and then execute the methodology themselves, often at minimal cost. This exploitation undermines the intellectual effort invested by research agencies and violates the ethical norms of professional collaboration.

3. **Skewed Study Populations:** Clients might direct researchers to target specific respondent groups that yield favorable outcomes. *For example, a retail company launching a new product may direct the researcher to conduct surveys only among loyal customers, rather than a representative sample of the target market.* This approach would inflate positive feedback, misrepresenting the product's actual market potential and biasing the results.

4. **Exceeding Research Scope:** Clients occasionally demand recommendations or interpretations beyond the study's original scope. *For instance, a client conducting a market study on the popularity of electric vehicles among urban consumers may later demand recommendations on government policy lobbying strategies, which falls outside the original scope of the research and lacks adequate exploratory support.*

   Researchers should clearly define objectives to limit the scope for manipulation or unethical interventions by clients. It is the researcher’s responsibility to educate clients about the importance of unbiased results and the risks of compromising data integrity. When faced with unethical demands, researchers should avoid formulating strategies or making recommendations based on compromised data and, if necessary, terminate the study to protect their professional integrity.

14.9.2 Ethical Code regarding Respondents

Respondents play a critical role in research, and their rights and dignity must be safeguarded. To prevent unethical practices, organizations like the American Association for Public Opinion Research have established guidelines emphasizing the ethical treatment of respondents.

Following are the key ethical principles regarding respondents:

1. **Avoiding Harm and Misuse:** Researchers must avoid practices that may harm, humiliate, or mislead respondents. Respondent information must remain confidential unless explicit consent is provided for disclosure or alternative uses.

2. **Study Disclosure:** Respondents must be fully informed about the purpose, nature, and requirements of the study, ensuring they understand the types of questions and the time commitment involved. For example, *in a workplace climate survey, employees should be informed if they will be asked to participate in both initial questionnaires and subsequent focus group discussions to provide deeper insights into the findings.*

3. **Voluntary Participation:** Participation must be voluntary, without coercion or undue influence. For example, *A respondent who feels uncomfortable discussing financial difficulties should not be pressured to provide detailed information about their personal finances during a study on economic behavior.*

4. **Sensitivity and Respect:** Sensitive topics, such as personal beliefs or behaviors, require empathetic handling and open-ended questioning. Researchers must respect respondents' boundaries and ensure transparency about the study's objectives. For example, Questions about personal relationships or family dynamics should be phrased carefully to ensure respondents feel respected and comfortable sharing their experiences.

5. **Experimentation and Minimal Risk:** Respondents involved in experimental studies must be informed about the process and any potential risks, no matter how minimal. The researcher must prioritize minimizing harm. *For example, A study testing a new fitness program should ensure participants are fully informed about potential physical strain or injury risks involved in the activities.*

6. **Consent and Agreement:** Researchers should obtain clear, informed consent from respondents, whether through written or verbal agreements, to avoid misunderstandings or legal disputes. *For example, observing a respondent's behavior during a personal care product study requires explicit consent to address potential privacy concerns*.

14.10 LET US SUM UP

This unit emphasizes the critical role of ethics in research, ensuring credibility, reliability, and fairness throughout the research process. Ethical

practices safeguard participant rights, maintain scientific integrity, and foster trust between researchers and stakeholders. Researchers are obligated to align their work with societal well-being while complying with legal and regulatory standards, such as GDPR, to protect sensitive data and ensure transparency.

The unit identifies key ethical issues in research, including informed consent, which ensures participants voluntarily partake in studies with a full understanding of their purpose, risks, and benefits. Anonymity and confidentiality are crucial for protecting participant identities and sensitive data, especially in studies involving workplace culture or customer behavior. While deception may sometimes be necessary to avoid biased results, it must be justified, limited, and followed by debriefing to mitigate harm. Plagiarism, another significant concern, undermines the integrity of research and can lead to severe academic and professional consequences.

Ethical guidelines for business research address responsibilities for decision-makers and respondents. Decision-makers are urged to avoid manipulating results, misusing proposals, or exceeding the scope of the research. For respondents, the focus is on safeguarding their dignity through voluntary participation, clear communication, and minimizing risks, particularly in studies involving sensitive topics like mental health or financial behavior.

Eventually research practices uphold the rights of participants, ensure data integrity, and enhance trust in the research process. By adhering to ethical guidelines, researchers contribute positively to knowledge while respecting the dignity of individuals and maintaining the credibility of organizations involved.

14.11 KEYWORDS

Anonymity: Ensuring that participants' identities remain unknown to both the researcher and any third parties, protecting their privacy and confidentiality.

Beneficence: The ethical principle of maximizing benefits while minimizing harm to research participants.

Confidentiality: The obligation of researchers to protect all information gathered from participants and ensure it is not disclosed to unauthorized individuals.

Conflict of Interest: A situation where a researcher's personal or professional interests may compromise the objectivity or integrity of their research.

Consent: The process by which participants voluntarily agree to take part in research after being fully informed about the study's purpose, risks, and benefits.

Deception: The intentional act of misleading participants about the true purpose or nature of a study to ensure unbiased results.

Debriefing: The process of providing participants with full disclosure about the research objectives and methods after their participation, particularly in cases where deception was involved.

Ethical Guidelines: Established principles and standards that guide researchers in conducting studies responsibly and ethically.

Informed Consent: A process that ensures participants are fully aware of all relevant aspects of a study, including risks and benefits, before agreeing to participate.

Plagiarism: The unethical use of someone else's work, ideas, or data without proper acknowledgment or permission.

Transparency: Openness and honesty in conducting and reporting research, ensuring that methods and findings are accessible and clear.

Voluntary Participation: The principle that participants should engage in research willingly, without coercion or undue pressure, and can withdraw at any time.

14.12 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress B:

5. (i) b (ii) c (iii) b (iv) b

14.13 TERMINAL QUESTIONS

1) What are the key ethical challenges associated with maintaining anonymity and confidentiality in research?
2) Explain the ethical concerns related to deception in research and the importance of debriefing.
3) How do ethical guidelines by organizations like the AAA and ICMR guide research involving vulnerable populations?
4) Examine the role of informed consent in ethical research and highlight historical cases where its absence led to ethical violations.
5) Critically discuss the ethical considerations involved in using deception as a research method, including the conditions under which it may be justified
7) Define ethical guidelines regarding decision makers
8) Define plagiarism and briefly explain its consequences.

14.14 FURTHER READINGS

1. "Research Ethics: A Handbook of Principles, Methods, and Cases" by C. R. Kothari and Gaurav Garg
2. "Research Ethics: A Handbook of Principles, Methods, and Cases" by William O. Jenkins

3. "Ethics in Research: A Global Perspective with Indian Context" by T. D. Ramabadran
4. "Ethical Issues in Research" by Bernard E. Rollin
5. "Ethical Guidelines for Social Science Research in India" by R. N. Sharma
6. "Research Ethics: A Global Perspective with Indian Context" by R. K. Sinha and Neelam Mehra

Note	These questions are helpful to understand this unit. Do efforts for writing the answer of these questions but do not send your answer to university. It is only for your practice.

UNIT 15 REPORT WRITING

Structure

15.0 OBJECTIVES

After studying this unit, you will be able to:

- Define a report;
- Explain the importance of reporting;
- Describe the subject matter of various types of reports;
- Identify the stages involved in preparing a report;
- Highlight the key characteristics of a good report;
- Explain the structure and components of a report; and
- Differentiate between a well-written and poorly written report.

15.1 INTRODUCTION

When you complete a business research project, the final step is **report writing**. This is where all the hard work of collecting, analyzing, and interpreting data comes together to communicate your findings. A report isn't just a formality—it's the way you share the results of your research in a clear

and structured way. Without a well-written report, the value of your research might not reach your intended audience.

In business research, **reporting** means sharing the facts and findings you've gathered through your research with key stakeholders, whether they are managers, investors, or business clients. You've collected data, analyzed it, and drawn conclusions—but now you need to communicate what you've discovered and what it means for the business.

Think of a report as more than just a document—it's an **account** that summarizes the most important details of your research. It includes what you did, the significant findings you uncovered, and what actions should be taken based on those findings. A report helps turn raw data into insights that can guide business decisions and improve outcomes.

For example, let's say you've conducted research on customer satisfaction for a retail brand. You might have numbers, statistics, and charts showing what customers like or dislike about the brand's services. However, those numbers alone don't tell the full story. A good report takes all that data, organizes it, explains what it means for the business, and makes actionable recommendations to improve customer satisfaction.

In this unit, we'll discuss the significance of business reports, explore the different types of reports you may encounter, and break down the steps to write them effectively. By the end of this unit, you'll be able to write a business research reports in clear, professional, and impactful manner.

15.2 SIGNIFICANCE OF A REPORT

A research report is a well-crafted document that outlines the processes, data, and findings of a systematic investigation. It is an important document that serves as a first-hand account of the research process, and it is typically considered an objective and accurate source of information.

Figure 15.1: Significance of a Research Report

In many ways, a research report can be considered as a summary of the research process that clearly highlights findings, recommendations, and other important details. Reading a well-written research report should provide you with all the information you need about the core areas of the research process.

A research report is crucial for several reasons:

1. **Communication of Findings**: It effectively communicates the results of research to a broader audience, including stakeholders, policymakers, and the academic community.

2. **Documentation**: It serves as a formal record of the research process, methodologies, and outcomes, which can be referenced in future studies or for replication purposes.

3. **Contribution to Knowledge**: Research reports add to the existing body of knowledge in a field, providing insights, data, and analysis that can inform future research, practice, and policy.

4. **Critical Evaluation**: They allow for the peer review and critique of research methods and findings, which is essential for validating the research and its contributions.

5. **Guiding Decision-Making**: In applied fields, research reports can inform decisions in business, healthcare, education, and other sectors, helping practitioners implement evidence-based strategies.

6. **Funding and Support**: A well-structured research report can attract funding and support for future research endeavors by demonstrating the significance and impact of the work.

7. **Ethical Accountability**: It ensures that researchers are accountable for their findings and methodologies, promoting transparency and ethical standards in research.

Overall, research reports play a vital role in advancing knowledge, fostering collaboration, and influencing practice across various domains.

15.3 TYPES OF REPORT

Reports can be broadly categorized into two main types:

1. **Oral or Verbal Report:** This involves presenting findings verbally in person. Examples include presenting at a conference, seminar, or reporting findings directly to superiors.

2. **Written Report:** Written reports are more formal, structured, and widely used in business research. These can take various forms, such as:

   ○ **Sentence Form Reports**: These are written in paragraphs, communicating the findings in detailed sentences.

   ○ **Tabular Reports**: These present data in tables, focusing on figures and numbers for clarity.

- **Graphic Reports**: Data is communicated through graphs, charts, and diagrams to illustrate trends and patterns visually.
- **Combined Reports**: These use a mix of the above methods—sentence form, tables, and graphics—to provide a comprehensive view of the research findings. This is often the most popular format for research reports.

Business reports can vary in length and form, depending on the purpose and context. For example:

- **Business organizations** often prefer short reports in letter form.
- **Banks and financial institutions** tend to use data-heavy reports in tables.
- **Government reports** are typically more comprehensive, covering detailed findings.
- **Research theses or dissertations** (such as those for a Ph.D.) are more elaborate, presenting in-depth analysis and methodology.

The results of a research inquiry can be presented in many different ways, depending on the target audience and purpose. These can include:

1. **Technical Report:** A technical report is a comprehensive and formal report that explains all aspects of the research. It is typically used for detailed evaluations, public dissemination, or record-keeping. Key features include:
   - Explanation of methodology
   - Objectives of the study
   - Assumptions or hypotheses
   - Data collection and analysis methods
   - Presentation of findings with supporting evidence
2. **Popular Report:** A popular report aims to communicate research findings to a general audience in an accessible and attractive manner. These reports use simple language, minimize technical jargon, and often feature graphs, charts, and visuals to keep the reader engaged. They focus on practical implications and are designed for a broader, non-expert audience.
3. **Research Article:** Research articles are shorter reports that summarize the findings of a study. These are commonly presented at seminars or published in academic journals. The goal is to contribute new knowledge to the field and get the research recognized.
4. **Monograph:** A monograph is an in-depth study or long essay on a single subject. It is often used to present a detailed analysis of a specific research topic.

Other classifications of reports can be based on the approach or the nature of presentation:

- **Journalistic Report**: Prepared by journalists, these reports are intended for media publication and contain news or information value.
- **Business Report**: A report focused on business activities and used for internal communication within an organization. It helps in decision-making and typically covers specific business areas.
- **Project Report**: A report detailing the progress, challenges, and results of a specific business project. It could be prepared by individuals or teams within a business.
- **Dissertation**: A lengthy and detailed report typically submitted as part of the requirements for higher academic degrees. It explores a research topic in great depth.
- **Enquiry Report (Commission Report)**: A report created by a commission or group of experts investigating a specific matter. These reports often contain recommendations and expert opinions to guide decision-making.

Each type of report serves a distinct purpose in business and research and can vary in terms of structure, detail, and audience. Knowing the right type to use for your research is essential to ensure it meets its intended goal effectively.

15.4 STAGES IN PREPARATION OF A REPORT

Research reports are the product of slow, painstaking and accurate work. Therefore, the preparation of the report may be viewed in the following major stages.

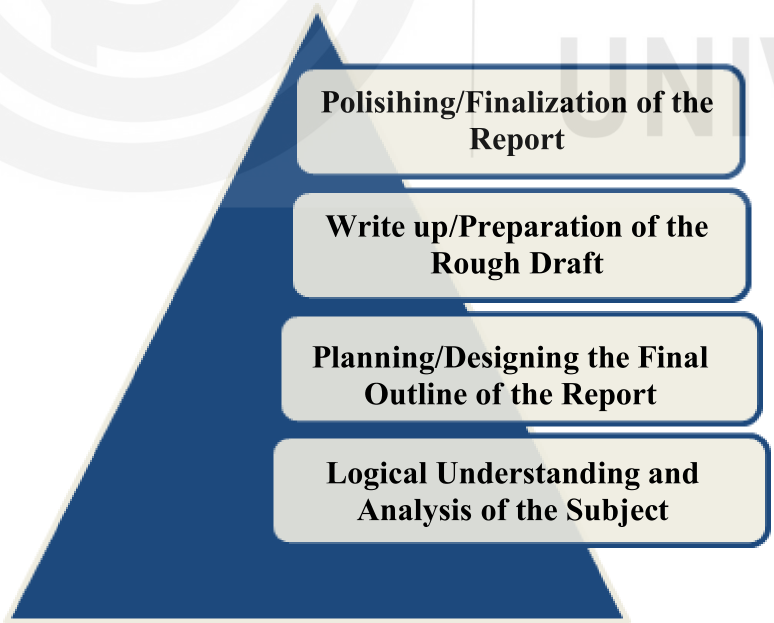


Figure 15.2: Major Stages in Preparation of the Report

1. **Logical Understanding and Analysis of the Subject Matter:** It is the first stage which is primarily concerned with the development of a subject. There are two ways to develop a subject viz. a. logically and b.

chronologically. The logical development is done on the basis of mental connections and associations between one aspect and another by means of logical analysis. Logical treatment often consists of developing material from the simple to the most complex.

Chronological development is based on a connection or sequence in time or happening of the events. The directions for doing something usually follow the chronological order.

2. **Planning/Designing the Final Outline of the Report:** It is the second stage in writing the report. Having understood the subject matter, the next stage is structuring the report and ordering the parts and sketching them. This stage can also be called the planning and organization stage. Ideas may pass through the author's mind. Unless he first makes his plan/sketch/design he will be unable to achieve a harmonious succession and will not even know where to begin and how to end. Better communication of research results is partly a matter of language but mostly a matter of planning and organizing the report.

3. **Write up/Preparation of the Rough Draft:** The third stage is the write up/drafting of the report. This is the most crucial stage to the researcher, as he/she now sits to write down what he/she has done in his/her research study and what and how he/she wants to communicate the same. Here the clarity in communicating/reporting is influenced by some factors such as who the readers are, how technical the problem is, the researcher's hold over the facts and techniques, the researcher's command over language (his communication skills), the data and completeness of his notes and documentation and the availability of analyzed results. Depending on the above factors some authors may be able to write the report with one or two drafts. Some people who have less command over language, no clarity about the problem and subject matter may take more time for drafting the report and have to prepare more drafts (first draft, second draft, third draft, fourth draft etc.,)

4. **Finalization of the Report:** This is the last stage, perhaps the most difficult stage of all formal writing. It is easy to build the structure, but it takes more time for polishing and giving finishing touches. Take example of the construction of a house. Up to the roofing (structure) stage the work is very quick but by the time the building is ready, it takes up a lot of time.

   The rough draft (whether it is second draft or 'n'th draft) has to be rewritten, polished in terms of requirements. The careful revision of the rough draft makes the difference between a mediocre and a good piece of writing. While polishing and finalizing one should check the report for its weaknesses in logical development of the subject and presentation cohesion. The researcher should also check the mechanics of writing — such as language, usage, grammar, spelling and punctuation.

15.5 CHARACTERISTICS OF A GOOD REPORT

A research report is a vital medium for communicating research findings. To ensure the report is effective, it should possess the following key characteristics:

1. **Clarity and Precision**: The report must clearly communicate the **what**, **why**, **who**, **whom**, **when**, **where**, and **how** of the research. It should be free from ambiguity, using familiar words and straightforward language. Avoid excessive jargon, wordiness, and complex phrasing that could confuse the reader.

2. **Appropriate Length and Readability**: The report should be of a length that sufficiently covers the subject matter but remains concise enough to sustain the reader's interest. It should be structured with proper paragraphing, short sentences, and visual aids like charts or graphs to enhance readability. Even technical content should be presented in an accessible way.

3. **Logical Organization and Coherence**: A good report must have a clear, logical structure with smooth transitions between sections. Each part should flow logically from the previous one, ensuring coherence throughout the report.

4. **Engagement and Interest**: A report should be engaging and not dull. It should maintain the reader's interest by presenting findings in a way that is both informative and thought-provoking, avoiding unnecessary complexity while highlighting key insights.

5. **Accuracy and Objectivity**: A report must be factually accurate, free from exaggerations, and written in an objective tone. It should present findings clearly and truthfully, avoiding subjective opinions or unsupported claims.

6. **Well-Defined Conclusions**: The report should draw clear and sound inferences based on the research data. Avoid repeating tables or figures verbatim in the text—focus on interpreting the data and drawing meaningful conclusions from it.

7. **Correct References and Bibliography**: Footnotes, citations, and the bibliography should follow the correct format and have to be complete. Proper referencing adds credibility and allows readers to trace the sources of information.

8. **Attractive and Error-Free Presentation**: The report should have a neat and clean appearance, whether typed or printed. It should be free from language errors, factual inaccuracies, spelling mistakes, and calculation errors. Attention to detail is essential to ensure professionalism and accuracy.

By adhering to these characteristics, a researcher can ensure that their report is effective, professional, and well-received by the intended audience.

Check Your Progress A:

1) What do you mean by a report?

....

....

....

....

....

2) What is the significance of a report?

....

....

....

....

....

3) What is a popular report?

....

....

....

....

....

4) What is meant by an article?

....

....

....

....

....

5) What do you mean by verbal reporting?

....

....

....

....

....

6) Choose the correct option

   i) What is the primary purpose of a business research report?

      a) to collect data without analysis

      b) to present research findings in a structured and clear manner

      c) to replace verbal communication in business settings

      d) to increase the length of research studies

   ii) Which of the following is a key stage in preparing a research report?

      a) skipping data verification

      b) avoiding peer review

      c) planning and designing the report outline

      d) writing without revising the draft

   iii) Which of the following is NOT a characteristic of good research report?

      a) clarity and precision

      b) logical organization and coherence

      c) excessive jargon and complex language

      d) well defined conclusions

15.6 STRUCTURE OF THE RESEARCH REPORT

Under this head, the format/outline/sketch of a comprehensive technical report or research report is discussed below.

STRUCTURE OF A RESEARCH REPORT

PREFATORY ITEMS

1. Blank sheet
2. Title page
3. Approval sheet (if any)
4. Rsearcher's declaration
5. Dedication (if any)
6. Preface and/or acknowledgments
7. Table of contents
8. List of tables
9. List of graphs/charts/figures
10. List of cases, if any
11. Abstract or high lights (optional)

TEXT OR BODY

Chapter 1: Introduction

Chapter 2 to n: Presentation & Description of Evidence

Chapter n+1: Summary Conclusion and Recommendations

TERMINAL ITEMS

1. Appendix, if any
2. Glossary, if any
3. Bibliography
4. Index
5. Blank sheet

Let us discuss these items one by one in detail.

A research report has a number of clearly defined sections. The headings of the sections and their order may differ from one situation to another. The contents of a report can broadly be divided into three parts as:

1) The front matter or prefatory items.
2) The body or text of the report.
3) The back matter or terminal items. The following chart summarizes the broad sequence of the contents of a research report.

15.7 PREFATORY ITEMS

The various preliminaries to be included in the front pages of the report are briefly narrated hereunder:

1) **Title Page:** The first page of the report is the title page. The title page should carry a concise and adequately descriptive title of the research study, the name of the author, the name of the institution to whom it is submitted, the date of presentation.

2) **Approval Sheet:** If a certificate of approval is required either from the research supervisor or from the institution which provided the research facilities, it must be given.

3) **Researcher's Declaration:** Generally, the researcher has to declare/ certify that it is his/her bonafide and original work done by him/her.

4) **Dedication:** If the author wants to dedicate the work to whomever he/she likes, he/she may do so.

5) **Preface or Acknowledgements:** A preface includes the background and reasons for the study. This is an appropriate place for him/her to make acknowledgements also. But if the researcher has opted to discuss the significance, reasons of the study elsewhere in the report he/she may not write 'preface'. But he/she may use the page for only acknowledgements. In acknowledgements the researcher acknowledges the assistance and support received from individuals and organizations in conducting the research. It is intended to express his/her gratitude to them.

6) **Table of Contents:** A table of contents gives an outline of the contents of the report. It contains a list of the chapters and their titles with page numbers. It facilitates easy location of topics in the report. The chapter headings may be typed with capital letters.

7) **List of Tables:** The researcher must have collected a lot of data and analyzed the same and presented it in the form of tables. These tables may be listed chapter wise and the list be presented with page numbers for easy location and reference.

8) **List of Graphs/Charts/Figures:** If there are many graphs and charts, they should also be listed with page numbers, after the list of tables separately.

9) **List of Cases/Exhibits:** If there are many cases/exhibits they should also be listed.

10) **Abstract:** An abstract is a summary of the report. It should be as brief as possible and run about one or two pages. It is placed at the prefactory part of the report so that a reader can get a quick overview of the report. It contains a brief and precise statement of the purpose and a bare summary of the findings or the results of the study.

15.8 BODY OF THE REPORT

After the preliminary items, the body of the report is presented. It is the major and main part of the report. It consists of the text and context chapters of the study. Normally the body may be divided into 3 (three) parts.

i) The introduction

ii) The description and discussion of evidence and findings

iii) The summary, conclusions and recommendations

i) Introduction

Generally, this is the first chapter in the body of the report. It is devoted to introducing the theoretical background of the problem and the methodology adopted for attacking the problem.

It may consist of the following aspects:

– Significance and justification of the topic.
– Theoretical background of the topic.
– Statement of the problem.
– Review of literature.
– Objectives of the study.
– Hypotheses to be tested.
– Definition of special terms, concepts and units of study.
– Scope of the study – geographical scope i.e. area/places to be covered, content, scope i.e., aspects to be included/excluded.
– Period of study i.e., reference period.
– Sources of data i.e., primary or secondary or both.
– Methods of data collection i.e., sample or census.
– Sampling design.
– Data collection instruments.
– Field work.
– Data processing and analysis plan.
– Limitations of the study, if any.
– An overview of the report i.e., chapter plan.

ii) Description and Discussion of Evidence

This is the major and main part of the report. It is divided into several chapters depending upon the number of objectives of the study, each being devoted to presenting the results pertaining to some aspect. The chapters should be well balanced, mutually related and arranged in logical sequence. The results should be reported as accurately and completely as possible explaining their bearing on the research questions and hypotheses.

Each chapter should be given an appropriate heading. Depending upon the need, a chapter may also be divided into sections. The entire verbal presentation should run in an independent stream and must be written according to best composition rules. Each chapter should end with a summary and lead into the next chapter with a smooth transition sentence.

While dealing with the subject matter of text the following aspects should be taken care of. They are:

1) Headings
2) Quotations
3) Footnotes
4) Exhibits

1) **Headings.** The following types of headings are commonly used.

   – CENTRE HEAD (All capitals, without underlining)
   – Centre Subhead (Capital and lower case, with underlining)
   – SIDE HEAD (All capitals without underlining)
   – Side Sub Head (Capital and lowercase letters with underlining)
   – Paragraph Head followed by a colon (Capital and lower-case underline)

Which combinations of headings to use depends on the number of classifications or divisions that a chapter has. The headings are illustrated below:

Centre Head. A Centre head is typed in all capital letters. If the title is long, the inverted pyramid style (i.e., the second line shorter than the first, the third line shorter than the second) is used. All caps headings are not underlined. Underlining is unnecessary because capital letters are enough to attract the reader's attention.

Example: CHALKING OUT A PROGRAMME FOR IMPORT
SUBSTITUTION AND EXPORT PROMOTION

Centre Subhead. The first letter of the first and the last word and all nouns, adjectives, verbs and adverbs in the title are capitalized. Articles, prepositions and conjunctions are not capitalized.

Example:

Chalking out a Programme for Import Substitution and Export Promotion

Side Heads. Words in the side head are either written in all capitals or capitalized as in the centre sub head and underlined.

Example: Import Substitution and Export Promotion

Paragraph Head. Words in a paragraph head are capitalized as in the centre sub head and underlined. At the end, a colon appears, and then the paragraph starts.

Example: Import Substitution and Export Promotion: The Seventh Five-Year Plan of India has attempted ……

2) **Quotations**

Quotation Marks: Double quotation marks (“ ”) are used. A quotation within a quotation is put in single quotation marks (‘ ’). Example: He said, “To the selfish, ‘freedom’ is synonymous with license”.

When to Use Quotation Marks: Quotation marks are used for

1) a directly quoted passage or word.
2) a word or phrase to be emphasized, and
3) Titles of articles, chapters, sections of a book, reports, and unpublished works.

How to Quote: a) All quotations should correspond exactly to the original in *wording, spelling, and punctuation.*

b) Quotations up to three typewritten lines are run into the text.

c) Direct quotations over three typewritten lines are set in indented paragraphs.

d) Quotation marks are not used for indented paragraphs.

Five ways of introducing a Quotation: These are given below.

a) **Introduction:** He/she said, “The primary test of success in a negotiation is the presence of goodwill on both sides”.

b) **Interpolation:** “The primary test of success in a negotiation”, he/she said, “is the presence of goodwill on both sides”.

c) **End Reference:** “The primary test of success in a negotiation is the presence of goodwill on both sides”, he/she said.

d) **Indented Paragraph:** He/she said: For the workers no real advance in their standard of living is possible without steady increase in productivity because any increase in wages generally, beyond certain narrow units, would otherwise be nullified by a rise in prices.

e) **Running into a Sentence:** He/she recommended that “joint management councils be set up in all establishments in the public as well as private sector in which conditions favourable to the success of the scheme exist”.

3) Footnotes

Types of Footnotes: A foot note either indicates the source of the reference or provides an explanation which is not important enough to include in the text.

In the traditional system, both kinds of footnotes are treated in the same form and are included either at the bottom of the page or at the end of the chapter or book.

In the modern system, explanatory footnotes are put at the bottom of the page and are linked with the text with a footnote number. But source references are incorporated within the text and are supplemented by a bibliographical note at the end of the chapter or book.

Rationale of Footnotes: Footnotes help the readers to check the accuracy of the interpretation of the source by going to the source if they want to. They are also an acknowledgement of the author's indebtedness to the sources. They lend authority to the work and help the readers to distinguish between the author's own contribution and that of others.

Where to put the Footnote: Footnotes appear at the bottom of the page or at the end of the chapter (before the appendices section).

Numbering of Footnotes:

a) For any editorial comment on the chapter or title, an asterisk is used.

b) In the text Arabic numerals are used for footnoting. Each new chapter begins with number 1.

c) The number is typed half a space above the line or within parentheses. No space is given between the number and the word. No punctuation mark is used after the number.

d) The number is placed at the end of a sentence or, if necessary to clarify the meaning, at the end of the relevant word or phrase. Commonly, the number appears after the last quotation mark. In an indented paragraph, the number appears at the end of the last sentence in the quotation.

4) Exhibits

Tables:

Reference and Interpretation: Before a table is introduced, it is referred to in the text (e.g., see Table 1.1; refer to Table 1.1; as in Table 1.1; Table 1.1 indicates). A table is meant only to expand, clarify, or give visual explanations rather than stand by itself. The text should highlight the table's focus and conclusions.

Identification:

a) Each table is given a number, title, and, if needed, a subtitle. All identifications are centred.

b) Arabic numerals, instead of Roman numerals or capital letters, are

recommended for numbering the tables. Usually, technical monographs and books contain many tables. As the number increases, Roman numerals become unfamiliar to the reader. Roman numerals also occupy more space than Arabic numerals. If there are more than 26 tables, capital letters will not be sufficient to identify them.

Tables can be numbered consecutively throughout the chapter as 1.1, 1.2, 1.3,… wherein the first number refers to the chapter and the second number to the table.

b) For the title and subtitle, all capital letters are used.

c) Abbreviations and symbols are not used in the title or subtitle.

Checklist: Relevance, accuracy, and clarity are of utmost importance in tables. When entering the table, check the following:

1) Has the explanation and reference to the table been given in the text?
2) Is it essential to have the table for clarity and extra information?
3) Is the representation of the data comprehensive and understandable?
4) Is the table number correct?
5) Are the title and subtitle clear and concise?
6) Are the column headings clearly classified?
7) Are the row captions clearly classified?
8) Are the data accurately entered and represented?
9) Are the totals and other computations correct?
10) Has the source been given?
11) Have all the uncommon abbreviations been spelt out?
12) Have all footnote entries been made?
13) If column rules are used, have all rules been properly drawn?

Illustrations: Illustrations cover charts, graphs, diagrams, and maps. Most of the instructions given for tables hold good for illustrations.

Identification: Illustrations are identified as FIGURE, CHART, MAP or DIAGRAM. The identification marks (i.e. number, title, and, if any, sub title) are put at the bottom, because an illustration, unlike a table, is studied from bottom upwards.

15.9 TERMINAL ITEMS USED IN REPORT WRITING

This section follows the text. First comes the appendices section, then the bibliography and glossary. Each section is separated by a divider page on which only the words APPENDICES, BIBLIOGRAPHY, or GLOSSARY all in capital letters appear.

All reference section pages are numbered in numerals in continuation with the page numbers of the text.

1) **Appendices**

What goes into an Appendix: a) Supplementary or secondary references are put in the appendices section. But all primary reference material of immediate importance to the reader is incorporated in the text. The appendices help the author to authenticate the thesis and help the reader to check the data.

b) The material that is usually put in the appendices is indicated below:

1) Original data
2) Long tables
3) Long quotations
4) Supportive legal decisions, laws and documents
5) Illustrative material
6) Extensive computations
7) Questionnaires and letters
8) Schedules or forms used in collecting data
9) Case studies / histories
10) Transcripts of interviews

Numbering of Appendices: The appendices can be serialized with capital letters (Appendix A, Appendix B) to differentiate from the chapter or table numbers.

References to Appendices:

a) In the text, the reader's attention is drawn to the appendices as in the case of tables.

b) All appendices are listed in the table of contents.

2) **Bibliographies Positioning of the Bibliography:** The bibliography comes after the appendices section and is separated from it by a division sheet written BIBLIOGRAPHY. It is listed as a major section in all capital letters in the table of contents.

A bibliography contains the source of every reference cited in the footnote and any other relevant works that the author has consulted. It gives the reader an idea of the literature available on the subject that has influenced or aided the author.

Bibliographical Information: The following information must be given for each bibliographical reference.

Books

1) Author(s)
2) Title (underlined)
3) Place of publication
4) Publisher
5) Date of publication
6) Date of issue

Magazines and Newspapers

1) Author(s)
2) Title of the article (Within quotation marks)
3) Title of the magazine (underlined)
4) Volume number (Roman numerals)
5) Serial number (Arabic numerals)

3) **Glossary**

What is a Glossary: A glossary is a short dictionary giving definitions and examples of terms and phrases which are technical, used in a special connotation by the author, unfamiliar to the reader, or foreign to the language in which the book is written. It is listed as a major section in capital letters in the table of contents.

Positioning of a Glossary: The glossary appears after the bibliography. It may also appear in the introductory pages of a book after the lists of tables and illustrations.

Order of Listing: Items are listed in alphabetical order.

Example:

Centre Heading is listed under C and not under H.

4) **Index**

Index may be either subject index or author index. Author index consists of important names of persons discussed in the report, arranged in alphabetical order. Subject index includes a detailed reference to all important matters discussed in the report such as places, events, definitions, concepts etc., and presented in alphabetical order. Index is not generally included in graduate / post graduate students research reports. However, if the report is prepared for publication or intended as a work of reference, an index is desirable.

15.10 CHECKLIST FOR THE REPORT

Figure.15.3: Checklist for the Report

Following are the main points of checklist which a researcher should keep in mind while finalising the report:

Title: Is the title clear and relevant to the report's content?

Objectives: Have the objectives of the report been clearly stated and met throughout the document?

Introduction: Does the introduction provide a good overview of the research topic and its significance?

Clarity of Language: Is the language simple, clear, and free of jargon? Is the report easy to understand for the target audience?

Logical Flow: Does the report follow a logical sequence? Is the information presented in a structured way, from introduction to conclusion?

Data Presentation: Are tables, charts, and graphs properly labelled and easy to interpret? Do they support the analysis in the report?

References and Citations: Are all sources of data and information correctly cited? Have proper referencing guidelines been followed?

Conclusions and Recommendations: Are the conclusions drawn from the research findings logical? Do the recommendations offer practical solutions or insights?

Grammar and Spelling: Has the report been proofread for grammatical errors, spelling mistakes, and punctuation issues?

Formatting: Is the report formatted according to the guidelines? Have headings, subheadings, font style, and size been consistently used?

Ethical Considerations: Does the report respect research ethics? Has consent been obtained where necessary? Is there any risk of plagiarism?

15.11 OPEN WRITING SOFTWARE

Open writing software refers to free and accessible tools that help in creating and editing documents without the need for expensive licenses. These tools are widely used in academic and professional writing for their cost-effectiveness and functionality.

Some popular open writing software includes:

- **Libre Office Writer**: A powerful free word processor, similar to Microsoft Word, that supports various file formats and offers strong formatting features.
- **Google Docs**: A web-based tool that allows real-time collaboration and document editing from any device.
- **Overleaf**: A LaTeX-based online editor, commonly used for writing scientific reports and research papers.
- **Apache Open Office Writer**: Another free word processor with features similar to LibreOffice, supporting multiple formats.

15.12 RESEARCH ETHICS

Research ethics refers to the set of principles that guide the conduct of researchers to ensure honesty, integrity, and respect for participants and data. Ethical considerations are crucial in maintaining the trustworthiness of the research and protecting the rights and well-being of all involved.

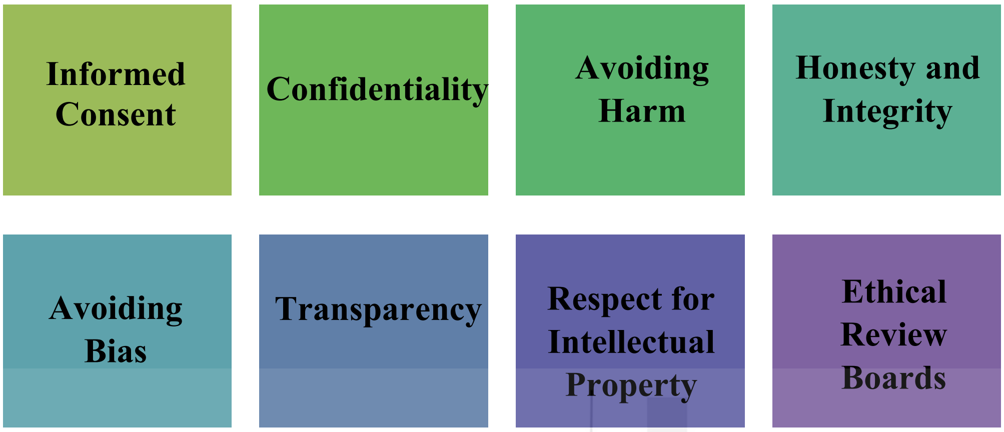


Figure 15.4: Key Aspects of Research Ethics

Key aspects of research ethics include:

1. **Informed Consent**: Before involving participants in any research, it is important to obtain their informed consent. This means participants should be fully informed about the purpose of the research, how the data will be used, and any potential risks involved. They should voluntarily agree to participate without any pressure or coercion.

2. **Confidentiality**: Protecting the privacy of participants is essential. Personal information and responses must be kept confidential and used only for the purposes agreed upon. Data should be securely stored to prevent unauthorized access.

3. **Avoiding Harm**: Researchers must ensure that their work does not cause any harm to participants, whether physical, emotional, or psychological. The well-being of participants should always be a priority.

4. **Honesty and Integrity**: Researchers are expected to conduct their work with honesty and integrity. This includes accurately reporting data, avoiding manipulation or fabrication of results, and giving proper credit to others' work through citations.

5. **Avoiding Bias**: Research should be objective and free from personal or external biases that could affect the results. Researchers should strive to remain neutral and not allow personal opinions or external pressures to influence the findings.

6. **Transparency**: Researchers should be open about the methods, data collection, and analysis procedures they use. This helps others replicate the study and ensures accountability in the research process.

7. **Respect for Intellectual Property**: It is important to respect the work of other researchers by citing sources appropriately and not using others' ideas without acknowledgment. Plagiarism is a serious ethical violation.

8. **Ethical Review Boards**: Many institutions require researchers to get approval from an ethical review board (or committee) before conducting research, especially if it involves human or animal subjects. This ensures that the research adheres to ethical guidelines and safeguards participants' rights.

15.13 PLAGIARISM

Plagiarism is the act of using someone else's work, ideas, or words without giving proper credit, and it is considered a serious ethical violation in research and academic writing. In simple terms, plagiarism is a form of intellectual theft and dishonesty.

There are different forms of plagiarism, including:

1. **Direct Plagiarism**: Copying someone else's words exactly as they appear without quotation marks or citations. This is the most obvious and severe form of plagiarism.
2. **Paraphrasing Plagiarism**: Rewriting someone else's ideas or text in your own words without giving proper credit. Even though the wording is different, the ideas are still not original.
3. **Self-Plagiarism**: Reusing your own previous work (such as assignments or papers) without acknowledging that it has been used before. This is often overlooked but is also a form of plagiarism.
4. **Mosaic Plagiarism**: Piecing together ideas, phrases, or sentences from different sources without proper citation, making it seem like your own work.
5. **Accidental Plagiarism**: Unintentionally failing to cite sources or misquoting them. Even if done unknowingly, it is still considered plagiarism.

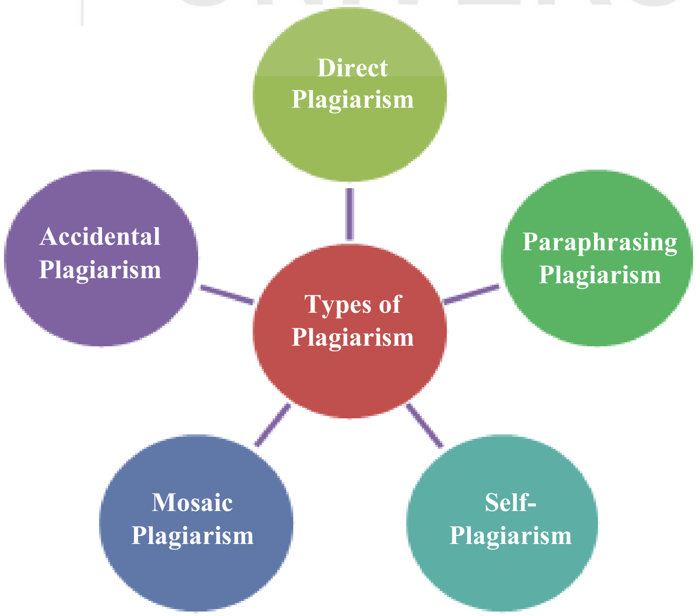


Figure 15.5: Types of Plagiarism

How to Avoid Plagiarism:

- **Cite Sources Properly**: Always give credit to the original author by citing their work, whether you are quoting directly or paraphrasing. Use the correct citation style as instructed (e.g., APA, MLA, Chicago).
- **Use Quotation Marks**: If you are using someone's exact words, make sure to place them in quotation marks and provide a reference to the original source.
- **Paraphrase Correctly**: When you rewrite someone else's ideas, ensure that the wording is completely your own and not too close to the original text. Still, provide a citation for the idea.
- **Reference All Sources**: Make sure to include all sources you have consulted, whether they are books, articles, websites, or other materials, in your reference list or bibliography.
- **Use Plagiarism Detection Tools**: Before submitting your report, use plagiarism detection software to check for unintentional similarities with other works. This helps ensure that your writing is original.

Check Your Progress B:

1) List the stages involved in the preparation of a report.

....

....

....

....

....

2) What are the ways of developing a subject?

....

....

....

....

....

....

3) What is meant by outlining the report?

....

....

....

....

....

....

4) Enumerate the characteristics of a good report.

....

....

....

....

....

5) Write short notes on the following:

   a) Plagiarism

   b) Research Ethics

   c) Open Writing Software

15.14 LET US SUM UP

The last step of the research process is writing a research report which acts as an informative account of the entire research study. Reporting can be done in two ways: oral or verbal and written. It can be in many forms; technical reports, popular reports, project reports, enquiry reports, articles or monographs. The process of writing a report includes comprehensible interpretation, planning, drawing the first outline of a report as well as refining/styling a report.

Research reports should be simple, brief, factual, rational, coherent and free from any complications. It must be short and precise, not containing value judgments and finally must be coherent and sequential. The look and feel of the document must be very simple: it is very important that this report follow the best rules of report writing.

A technical report is a detailed description of a study conducted accompanied by the front, the body and the back working. It involves the formation of an introduction, evidence part, the finding session, conclusions and the recommendation part. For the purpose of the text organization, the material is divided into appendices, bibliography, and a glossary.

Ethical considerations in research embrace the need to be truthful, and provide subject and data respect. They include informed consent, confidentiality, do not harm, do not mislead, free from opinion bias, disclosure versus secrecy, respect researcher ownership and drawing from ethical review boards. Conducting research and writing is forbidden by plagiarism a common violation of the ethical norms of obtaining new knowledge, and responding to it is possible when using quotation marks, paraphrasing the material, indicating all sources used in writing, as well as using specialized services for detecting cases of plagiarism.

15.15 KEYWORDS

Abstract: An abstract is a short summary of the report.

Article: A short paper prepared for publication in a journal/for presentation in a seminar/conference.

Bibliography: It is the list of all published and unpublished references used in the report arranged in alphabetical order.

Dissertation: A formal and lengthy discourse.

Footnote: It is an explanatory note/material source, given at the bottom of the page.

Glossary: A list of words.

Lay out: Sketch, design, structure.

Monograph: A treatise on a single subject.

Plagiarism: Plagiarism is the unethical practice of using someone else's work without proper credit, a serious ethical violation in research and academic writing.

Report: A report is an account of the research study.

Reporting: Reporting means communicating through a report.

Research Ethics: Research ethics is a set of principles guiding researchers to ensure honesty, integrity, and respect for participants and data.

Thesis: A formal and lengthy research report presented as part of the requirements for a degree.

15.16 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress A

6. (i) b (ii) c (iii) c

15.17 TERMINAL QUESTIONS

1) What is a report? What are the characteristics/qualities of a good report?
2) Distinguish between oral reporting and written reporting.
3) Differentiate between a technical report and a popular report.
4) What are the items that can be included in the Appendix?
5) What is reporting? What are the different stages in the preparation of a report?
6) What are the various aspects that have to be checked before going for final typing of the report?
7) What are the points to be kept in mind in revising the draft report?
8) Write short notes on:
   a) Characteristics of a good report.
   b) Prefatory items in a report
   c) Bibliography
   d) Research Article

15.18 FURTHER READINGS

1. V.P. Michael, Research Methodology in Management, Himalaya Publishing House, Bombay.
2. O.R. Krishna Swamy, Methodology of Research in Social Sciences, Himalaya Publishing House, Mumbai.
3. C.R. Kothari, Research Methodology, Wiley Eastern, New Delhi
4. Berenson, Conrad and Raymond Cotton, Research and Report Writing for Business and Economics, Random House, New York.

Note	These questions are helpful to understand this unit. Do efforts for writing the answer of these questions but do not send your answer to university. It is only for your practice.

ignou
THE PEOPLE'S
UNIVERSITY

Indira Gandhi National Open University
School of Management Studies